Département de Physique
Université de Fribourg
Suisse

Thèse

Applications of statistical physics to complex systems

présentée à la Faculté des Sciences
de l'Université de Fribourg (Suisse)
pour l'obtention du grade de
Doctor rerum naturalium

Matúš Medo
de
Slovaquie

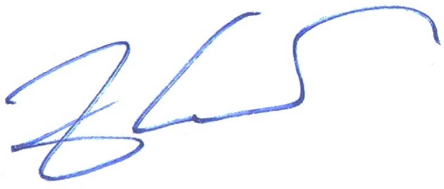
Thèse No. 1614
Imprimerie UniPrint
2008

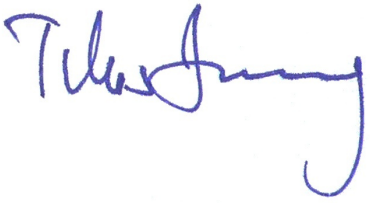
Acceptée par la Faculté des Sciences de l'Université de Fribourg (Suisse) sur la proposition de Dr. František Slanina et Prof. Vittorio Loreto (experts), Prof. Yi-Cheng Zhang (directeur de thèse) et Prof. Dionys Baeriswyl (président du jury).

Fribourg, le 17 septembre 2008

Le Directeur de thèse: **Le Doyen:**

Prof. Yi-Cheng Zhang Prof. Titus Jenny





To my wife and to my parents.

Abstract

Knowledge and information make our lives easier and more enjoyable. In this thesis we use techniques and concepts from statistical physics and probability theory to explore several models of complex systems where information plays a prominent role. When building these models we tried to capture main aspects of real systems but also to keep the models simple and analytically solvable.

We begin the thesis with a short introduction to the field of complexity in Chapter 1. Trying to be brief, we only mention basic notions in the field and point to the relevant literature. Recommender systems serve to extract useful information from the expressed opinions of users about a particular set of objects. We explore them in detail in Chapter 2, where we describe two novel recommendation methods and compare them to other standard methods. In a market, the vendor has limited information about preferences of the buyers and the buyers have limited information about qualities (and other properties) of the offered products. Focusing on these information deficiencies, in Chapter 3 we present two complementary models which, while highly simplified, are able to capture many aspects of marketing (such as product differentiation, competition of vendors, partial information asymmetry, etc.). History of stock prices or rules of a game of chance also represent information which, when used properly, may allow one to make a profitable investment. Concerned mainly with the Kelly approach which is based on information theory, in Chapter 4 we investigate the problem of portfolio choice and the role of (limited) information. Finally, in Chapter 5 we provide a broader view on this thesis and on the science of complexity in general.

Resumé

La connaissance et l'information rendent notre vie plus facile et plus agréable. Dans cette thèse nous utilisons des techniques et des concepts de la physique statistique et de la théorie des probabilités afin d'explorer plusieurs modèles de systèmes complexes où l'information joue un rôle important. Lors de la construction de ces modèles, nous avons essayé de capturer les aspects principaux des systèmes réels, mais aussi de garder les modèles simples et analytiquement solubles.

Nous commençons la thèse avec une brève introduction aux systèmes complexes dans le Chapitre 1. Tout en essayant d'être bref, nous introduisons les notions de base de ce domaine et la littérature pertinente. Les "systèmes de recommandation" servent à extraire les informations utiles à partir des opinions exprimées par les utilisateurs sur un ensemble d'objets. Nous explorons ces systèmes dans le Chapitre 2, où nous proposons deux nouvelles méthodes de recommandation que nous comparons aux méthodes standards. Dans un marché, le vendeur dispose d'informations incomplètes sur les préférences des acheteurs et de même les acheteurs disposent d'informations incomplètes sur la qualité (et d'autres propriétés) des produits proposés. En mettant l'accent sur ces manques d'information, nous présentons dans le Chapitre 3 deux modèles complémentaires qui, bien que très simplifiés, sont capables de capter de nombreux aspects de la commercialisation (tels que la différenciation des produits, la concurrence entre vendeurs, l'asymétrie partielle de l'information, et ainsi de suite). L'histoire de l'évolution du prix des actions ou les règles d'un jeu de hasard représentent aussi des informations qui, lorsqu'elle sont utilisées correctement, peuvent permettre de faire un investissement rentable. En nous préoccupant principalement de l'approche de Kelly qui est basée sur la théorie de l'information, nous étudions dans le Chapitre 4 le problème de choix de portefeuille et le rôle de l'information limitée. Enfin, nous donnons dans le Chapitre 5 une vue d'ensemble sur cette thèse et sur la science des systèmes complexes en général.

Publications

1. ⋆ Matúš Medo, Distance-dependent connectivity: Yet Another Approach to the Small World Phenomenon, *Physica A* **360/2**, 617–628, 2006

2. Tao Zhou, Jie Ren, Matúš Medo, Yi-Cheng Zhang, Bipartite network projection and personal recommendation, *Phys. Rev. E* **76**, 046115, 2007

3. Yi-Cheng Zhang, Matúš Medo, Jie Ren, Tao Zhou, Tao Li, Fan Yang, Recommendation model based on opinion diffusion, *Europhysics Letters* **80**, 68003, 2007

4. Matúš Medo, Yi-Cheng Zhang, Market Model with Heterogeneous Buyers, *Physica A* **387**, 2889–2908, 2008

5. ⋆ Matúš Medo, Ján Smrek, Heterogeneous network with distance dependent connectivity, *EPJ B* **63**, 273–278, 2008

6. Matúš Medo, Yury M. Pis'mak, Yi-Cheng Zhang, Diversification and limited information in the Kelly game, *Physica A* **387**, 6151–6158, 2008

7. Linyuan Lü, Matus Medo, Yi-Cheng Zhang, Damien Challet, Emergence of product differentiation from consumer heterogeneity and asymmetric information, *EPJ B* **64**, 293–300, 2008

8. Paolo Laureti, Matúš Medo, Yi-Cheng Zhang, *Analysis of Kelly-optimal portfolios*, arXiv:0712.2771 (submitted to Quantitative Finance)

9. Matúš Medo, Chi Ho Yeung, Yi-Cheng Zhang, *How to quantify the influence of correlations on investment diversification*, arXiv: 0805.3397 (submitted to International Review of Financial Analysis)

Remark: Publications marked with ⋆ are not directly related to this thesis and my PhD study in Fribourg.

Acknowledgments

This dissertation would not have been possible without help and support of my supervisor, colleagues, friends, and family. First of all I would like to thank to my supervisor Yi-Cheng Zhang who has been a constant source of ideas, inspiration, and motivation. Other members of the theoretical physics section have been also helpful and made Fribourg a nice place to stay: Dionys Baeriswyl, Xavier Bagnoud, David Eichenberger, Bruno Gut, Peter Barmettler, Damien Challet, Joe Wakeling, Tao Zhou, Zike Zhang, Jianguo Liu, and also those who already moved to another nice places: Ying Jiang, Paolo Laureti, Marcel Blattner, and Lionel Moret. Without our great secretaries, Anne Fessler and Eliane Esseiva in particular, I would have got lost in Swiss administrative and regulations.

I feel indebted to all my collaborators who helped me to survive in the hard struggle which some call research. To be specific, results presented in Chapter 2 were obtained in collaboration with Tao Zhou, Jie Ren, Tao Li, and Fan Yang, results presented in Chapter 3 with Linyuan Lü and Damien Challet, and results presented in Chapter 4 with Paolo Laureti, Chi Ho Yeung, and Yury M. Pis'mak.

I cannot neglect those who I met back in Bratislava, in particular František Slanina, Vlado Černý, Martin Mojžiš, Ján Pišút, Ján Boďa, who greatly contributed to my understanding of science, and innumerable friends who made my time in Slovakia really great.

Above all, I thank to my wife for her love and to our families for their encouragement and support.

Contents

1 Complex systems, information, and physics

I think the next century will be the century of complexity.
Stephen Hawking

Physics—raised by Newton, Laplace, Maxwell, and others—was long considered as a search for fundamental laws. It was assumed that once these are discovered, everything will follow from them and the history of physics will come to its end. Physicists were dissecting the investigated systems, studying the resulting fragments, and hoping that later someone will come who will put the pieces together. Complex behavior, when encountered, was considered to be a result of our ignorance or our lack of control of the variables involved [1]. Admittedly, this reductionist approach has yielded many admirable achievements ranging from celestial mechanics to quantum theories. Yet, the first cracks have appeared on it with the advent of chaos theory [2–4] which learned us that simple rules can produce intrinsically complicated behavior and small changes can lead to large effects.

As interesting as chaos theory is, it seems to be separated from real world phenomena (with weather as one notable exception). To put it straight, chaos seems more as a mathematical curiosity with profound fundamental consequences than a useful tool providing useful implications for real life. Indeed, systems around us are often complicated and rich but scarcely they are chaotic in the mathematical sense. Does it mean that in fact reductionism is enough? Not really, on the halfway between simplicity and chaos there is something else—complexity [5].

1.1 Complexity and complex systems

Derived from the Latin word *complexus* (twisted together), complexity is well captured by the famous phrase of Phil Anderson: "More is different." To find a less metaphoric explanation is difficult—complexity is a young field and even the basic term "complex system" is not well defined yet. To obtain a better idea it's useful to think of systems which are not complex as in [5]. The classical me-

chanic system, pendulum, is simple because it can be understood as a system with one component (atoms constituting the pendulum are unimportant for its behavior). When secondary effects are present (*e.g.*, damping due to air friction, torque due to the Coriolis force), they can be considered as small perturbations and they do not give rise to new features. Another step could be a snowflake which looks complicated and can be explained only when interactions of innumerable water molecules are considered. And yet, it's not a complex system because it's too rigid, too ordered. One more step and there are chaotic systems (such as the logistic map or the Lorentz attractor) without order, organization, hierarchies, and stability. Hence, only when order and chaos are in balance we say that a system is complex.

To be more illustrative, here are a few examples of complex systems.

- Cellular automata (with "Game of life" as a prominent example [6]).
- Swarm intelligence (systems where decentralized local interactions give rise to self-organization) [7].
- Cells of living organisms (with complicated signaling and metabolic pathways) [8].
- The brain (a significant example of an adaptive complex system) [9].
- Epidemic spreading (which is similar to the percolation problem known from statistical physics) [10].
- Human society, economy, and markets [11, 12].

Most of these systems belong to a specific class of "complex adaptive systems". By the word adaptive we mean that parts of the systems have the capacity to change and learn from experience. The complex systems mentioned above have an important common feature: they involve a mix of positive and negative effects mixed which yield, when acting together, complex patterns and behavior. This idea is well illustrated on the example provided by Waldrop in [5]. When water is spilled over a table, it forms complicated shapes. These shapes result from two opposite forces—gravitation which tries to spread water wide and surface tension which tries to hold drops together. With only one of the two forces present, the resulting shapes would be much simpler (a thin one-molecule layer without surface tension or a one big drop without gravitation).[1] Notably, such a mixture of opposing effects is investigated in our market model in Sec. 3.1 where production of high quality goods lowers the vendor's profits and production of low quality goods lowers the vendor's sales.

[1] For a statistical physicist, a more familiar example is the Ising model where spin-spin interactions and heat fluctuations oppose each other. In this context, Philip Ball remarks that "A phase transition arises from compromises." [12].

To provide an opposing force to the previous specific paragraph, we list some general definitions of complex systems which were presented in the special edition of Science[2] devoted to complexity.

- A complex system is a highly structured system, which shows structure with variations (Goldenfeld and Kadanoff).
- A complex system is one whose evolution is very sensitive to initial conditions or to small perturbations, one in which the number of independent interacting components is large, or one in which there are multiple pathways by which the system can evolve (Whitesides and Ismagilov).
- A complex system is one that by design or function or both is difficult to understand and verify (Weng, Bhalla, and Iyengar).
- A complex system is one in which there are multiple interactions between many different components (Rind).
- Complex systems are systems in process that constantly evolve and unfold over time (Brian Arthur).

Of course, attempts to cover the topic in a few pages are vain, an interested reader can find more information in books [5, 11, 13–15]. For a brief overview of complex systems see [16], for an early discussion see [17]. The science of complexity is now mature enough to yield real life applications, for a collection of examples see [18].

1.1.1 Algorithmic complexity

Algorithmic complexity (often called Kolmogorov complexity, descriptive complexity, etc.) is a measure of the computational resources needed to reproduce a given object (usually data). For example, the string "aaaaaaaaaa" can be reproduced by the single program "print 'a' ten times" while the string "4a0WUdf9as" has no obvious short description. Hence the algorithmic complexity is small for ordered objects and large for objects with a high degree of randomness. In particular, for strings of a certain length, the algorithmic complexity is maximized by a string of random symbols (within the given character table). This indicates that the complex behavior mentioned above, characterized by the presence of long-range coherence or patterns, is not described well by algorithmic complexity. Simply, there are various notions of complexity and one has to be careful when moving from one field to another.

It has been proven (independently by Solomonoff and Kolmogorov) that as soon as one restricts ourself to *universal languages* (those in which a universal

[2] *Science* **284**, No. 5411, 1999.

Turing machine can be implemented), the choice of programming language is asymptotically irrelevant. Yet there is another hitch—algorithmic complexity is uncomputable. That means, there is no computer program that for any given data finds the shortest program that produces the data. Even worse, it is in general impossible to prove that a given program is the shortest one. The usual way to overcome this obstacle is to use other description methods than general-purpose universal languages mentioned above. Of course, there is a price one has to pay for this simplification—for any more specific description method there are some regularities in the data that become unnoticed.

1.1.2 Complex networks

In mathematics, graphs are investigated for centuries. However, it was not until 1959 when Hungarian mathematicians Erdös and Rényi devised the probabilistic model of a graph where edges are drawn with a certain probability p [19]. They generalized graphs from static objects for which exact mathematical theorems can be proven to complicated structures for which the typical behavior can be investigated. A real boom of interest in random graphs has come only after two groundbreaking papers [20, 21] which provided not only interesting mathematical models to study but also a connection to real systems which can be described by these models. Nowadays, complex networks represent a very active subfield of complex science. Research of networks focuses on understanding of their topological properties, empirical analyses, and designing of new models. For excellent reviews of the field see [22–26].

Among the most important properties of complex networks is the small-world effect (the shortest path between nodes of the network is short and scales with the logarithm of the network size), high clustering coefficient (if node A is connected with B and C, then B and C are connected with an excessive probability), scale-free degree distribution (the probability that a node has degree at least k decays as $k^{-\alpha}$), and community structure (nodes of a network can be divided into several groups which are strongly intraconnected and weakly interconnected). Various processes on networks—for example percolation and epidemiological processes, cellular automata, voter models, and diffusion—are also intensively studied [27]. The last mentioned, diffusion on networks, is in our focus in Secs. 2.2 and 2.4.

1.1.3 Power laws

The abundance of power laws in nature was known long before the complex science was established. Perhaps the first observation of a power law dates back to 1896 when the Italian economist Vilfredo Pareto in his book *Cours d'Economie Politique* made the striking remark "in all countries and at all times the extreme

distribution of income and wealth follows a power law behavior". By the power law behavior of extreme values is meant that the distribution decays as an inverse power of those values, $P(x) \sim x^{-\alpha}$ for x large (for obvious reasons, the name Pareto distribution is often used). Since Pareto's investigation of income and wealth distributions, power laws have been found in many other fields—they describe city populations, sizes of earthquakes, sizes of Moon craters, frequencies of word usages, firm sizes, numbers of scientific citations, and many other quantities [28–30].

Power law distributions have several peculiar features. The three basic statistical quantities of a power-law distribution (mean, median, and mode) can differ substantially, for $\alpha < 1$ the mean is not even defined. By the scale transformation $x \mapsto cx$ they preserve its functional form because $P(cx) = c^{-\alpha}P(x)$. This property led to one more name of power law distributions—they are often labeled as scale-free. The lack of a typical scale is important because it allows exceptional events to occur. While rare, they can have a sever impact on the system behavior. As an example of an extreme event we can take the 22% drop of the DJIA stock index in October 1987. If index returns would be governed by the normal distribution, this 20-sigma event (the standard deviation of the DJIA daily returns is roughly 1.1%) would have had the utterly negligible probability of the order of 10^{-176} which, with respect to the age of the universe, means that the event is impossible. Thus, scale-free and other broad distributions are necessary to capture extreme behavior of systems. For a well-written book about financial markets with strong emphasis on concepts of statistical physics see [31].

The ubiquity of power laws is due to the fact that they result from many theoretical models [32].[3] In particular, power laws are often linked to critical phenomena and complex systems [15].

1.2 Entropy and information

Entropy was originally introduced as a state function of a thermodynamic system. It is defined by the relation $\Delta S = \Delta Q/T$ and when multiplied by the system temperature T, it can be understood as the amount of the system energy which cannot be used to do thermodynamic work. Later in statistical mechanics, entropy was linked to the number of possible microstates Ω compatible with macroscopic properties of the system by the relation $S = k_B \ln \Omega$ and to the probabilities of various microstates by the relation $S = -k_B \sum_i p_i \ln p_i$. Later on it was shown that the principle of maximum entropy can be understood as the footstone of statistical mechanics [33, 34].

[3] A less noble reason for the ubiquity of power laws is that empirical data are often misinterpreted and a power law behavior is more a desire than a fact [30, 32].

Meanwhile, information theory went its own way, arriving at Shannon entropy which quantifies the minimum message length per symbol by the equation $H = -\sum_i p_i \log_2 p_i$ (here p_i is the probability of symbol i, binary coding of symbols is assumed and thus the unit of H is *bit*) [35]. It can be shown that this is the only formula (up to multiplication by a positive number) which is simultaneously: continuous (with respect to small changes of the probabilities), symmetric (with respect to reordering of possible outcomes), additive (*i.e.*, independence on how a process is divided into parts), and maximized by evenly distributed probabilities. Shannon entropy is often interpreted as a measure of uncertainty: when tossing a coin where only two outcomes are possible, the uncertainty is maximized for a fair coin ($p_1 = p_2 = 1/2$).

Closely related to Shannon entropy is the principle of minimum description length [36] which is sometimes interpreted as the essence of science [34]. This principle formalizes the famous Occam's razor by searching for the hypothesis that achieves the best compression of the data (thus, preferred are simple theories which explain a wide range of phenomena). The resulting tradeoff between the complexity of the hypothesis and the complexity of the data given the hypothesis is a natural barrier to overfitting. For an excellent and comprehensive treatment of relations between probability theory and science see [37].

As we have seen, entropy in physics and entropy in information theory are interconnected and both constitute bases in their fields. Furthermore, Shannon's work on information theory was later extended to the financial analysis by his PhD student J. L. Kelly. This extension, commonly known under the name "Kelly game", we investigate in detail in Sec. 4. In this thesis we usually use the word information in a looser sense without referring explicitely to entropy, probabilities, and coding.

2 Physicists approach to recommender systems

A wealth of information creates a poverty of attention.
Herbert Simon

The exponential growth of the Internet [38] and the World-Wide-Web [39] confronts us with the information overload: we face far too many data and data sources, making us unable to find the relevant results. As a consequence we need automated ways to deal with the data, to filter out the information relevant for us. One of the landmarks of information filtering is the advent of Internet search engines [40, 41]. However, they cannot fully solve the problem of information overload because they are not personalized—with the same query, different users receive identical answers. When historical track of users' activities is available, personalized recommendation is likely to produce better results than "recommendation for general audience". Accordingly, personal recommendation is often used in real online systems: music-sharing websites (`www.last.fm`), book-sharing websites (`www.booklamp.org`), online shops (`www.amazon.com`), social bookmarking (`www.CiteULike.com`), restaurants recommendation (`foodio54.com`), and so forth. The Internet supplemented with sophisticated social and recommendation tools has the potential to turn the world into a real global village [42] where space and time barriers do not exist anymore.

The importance of recommender systems for society and economy has resulted in high activity in this research field with contributions from marketing practice [43, 44], mathematical analysis [45], engineering science [46–48], but also from physics community [49–52]. Among the main directions of research we can mention correlation-based methods [53–55] and spectral methods [56–59]. The field is developing rapidly [60]. For example, only recently it was realised that the recommendation accuracy is not sufficient to quantify performance of recommender systems. Instead, one should assess the full list of items recommended to a given user, with the aim of maximising its usefulness as perceived by the user [61–63]. Notably, focus on usefulness is a common practice in commercial recommendation services (Google, for example) and represents the major challenge also for academia.

As a prelude to the discussion of recommender systems, in the following section we analyse a large dataset containing ratings from almost half million users for more than 17 000 movies. In Sec. 2.2 we describe a new recommendation method which is based on an analogy with the standard physics phenomenon, diffusion. In Sec. 2.3 we compare performances of several recommendation methods on various datasets. Finally in Sec. 2.4 we generalize the diffusion recommendation method so that it can be used on unweighted data.

2.1 A glance at the users' rating patterns

Testing datasets containing reliable user ratings are essential for development of recommendation methods. For a long time, the standard dataset used in most scientific papers was the one prepared by the lab GroupLens.[1] It contains approximately one million ratings from 6 040 users for 3 706 movies. At the end of 2007, the DVD rental company Netflix has started a million dollar competition, the Netflix Prize, for the best recommendation algorithm.[2] For the competition, a huge testing dataset containing more than 100 million ratings from 480 189 users for 17 770 movies was released (the integer scale $1, 2, 3, 4, 5$ is used). This dataset has one important advantage—it was generated using opinions given by Netflix customers. This makes the data more reliable because they are produced by real people shortly after watching the movies. In addition, these people have paid to see the movies and are interested to receive good recommendations for future movies which will be also paid.

To gain intuition about users' rating patterns, we first briefly overview basic statistical properties of the Netflix dataset. In Fig. 2.1 we investigate how many ratings individual users have given (left) and individual movies have received (right). As can be seen, both distributions are very broad. This means that while the average number of rating per user and per movie are around 200 and $5\,600$ respectively, there are many users who have given thousands ratings (movie addicts) and several movies who have received hundreds thousands ratings (so-called blockbusters). Heavy users are dangerous for two different reasons. First, this minority of users can considerably influence the resulting recommendations for all. Second, extreme users often turn out to be spammers trying to influence the results in favor of a particular object. Spamming recommender systems is a serious problem nowadays, we would like to investigate sensitivity of various recommendation methods to spamming in future.

Rating patterns of users and movies can be quantified by computing their averages and standard deviations, in Figs. 2.2 and 2.3 we show histograms of these quantities. As can be seen, both users and movies are very heterogeneous, differing largely both in the average ratings and in the dispersion of ratings.

[1] See `www.grouplens.org`.
[2] See `www.netflixprize.com`.

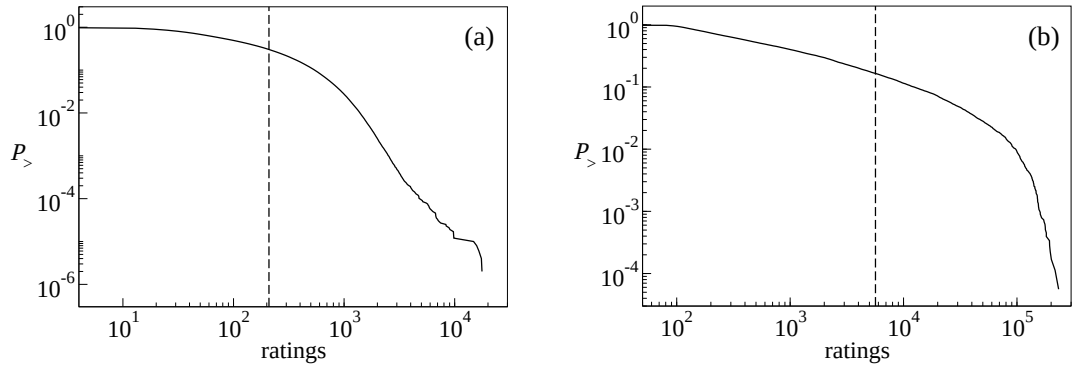


Figure 2.1: Distribution of given/received ratings for users (a) and movies (b). The average numbers of ratings are indicated by vertical dashed lines.

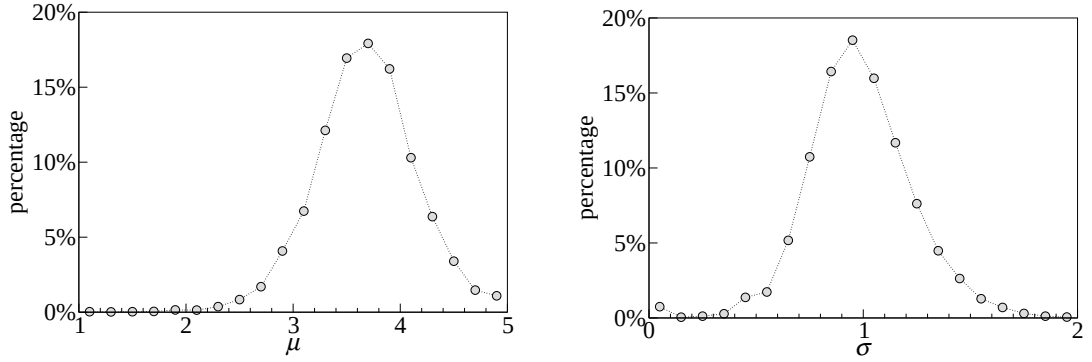


Figure 2.2: Histogram of rating averages and standard deviations for users in the Netflix dataset.

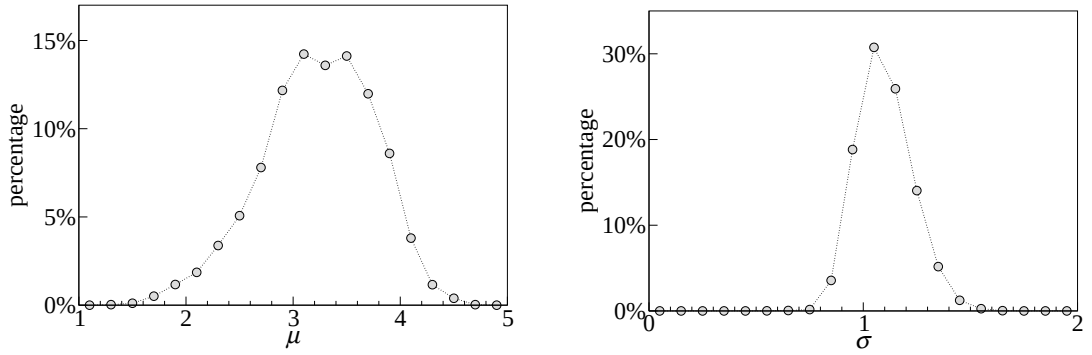


Figure 2.3: Histogram of rating averages and standard deviations for movies in the Netflix dataset.

2.2 Opinion diffusion: a new recommendation method

The danger is that it's all too easy to find apparent patterns in what's really random noise. If you use these mathematical hallucinations to predict ratings, you fail.
Wired magazine on the Netflix prize

In this section we describe a new recommendation method which is motivated by the classical physical phenomenon: diffusion. This method can be used for any data where users evaluate objects on an integer scale. Using data from a real recommender application (GroupLens project) we show that the presented model performs better better than the standard recommendation methods (for a more detailed comparison see Sec. 2.3. A Green function method is proposed to further reduce computation in some cases.

2.2.1 The model

In the input data, the total number of users we label as M and the total number of objects as N (since we focus here on the movie recommendation, instead of the general term *object* we often use the term *movie*). To make a better distinction between these two groups, for user-related indices we use lower case letters $i, j, k, \ldots$ and for movie-related indices we use Greek letters $\alpha, \beta, \gamma, \ldots$. We assume that users' assessments are given in the integer scale from 1 (very bad) to 5 (very good). The rating of user i for movie α we denote $v_{i\alpha}$. The number of movies rated by user i we label k_i. The rating data can be described by the weighted bipartite graph where the link between user i and movie α is formed when user i has already rated movie α and the link weight is $v_{i\alpha}$. Such a bipartite graph can give rise to two different types of graphs (often called *projections*): object-to-object and user-to-user. A general discussion on information networks can be found in [24], projections of bipartite graphs are closely investigated in [64, 65].

The recommendation process starts with preparation of a particular object-to-object projection of the input data. Projections usually lead to a loss of information. In order to eliminate this phenomenon, instead of merely creating a link between two movies, we link the ratings given to this pair of movies. As a result we obtain 25 separate connections (channels) for each movie pair. This is illustrated in Fig. 2.4 on an example of a user who has rated three movies; as a result, three links are created between the given movies. When we process data from all users, contributions from all users shall accumulate to obtain an aggregate representation of the input data: a weighted movie-to-movie network. From the methodological point of view, this model is similar to the well-known

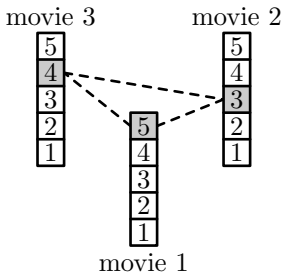


Figure 2.4: Graphical representation of the links created by a user who has rated only movies 1 (rating 5), 2 (rating 3), and 3 (rating 4).

Quantum Diffusion process (see [66, 67]).

To each user we need to assign a weight. In general, if user i has rated k_i movies, $k_i(k_i-1)/2$ links in the network are created (or fortified). If we set the user weight to $1/(k_i-1)$, the total contribution of user i is directly proportional to k_i, and this is a plausible premise.[3] Since the users who have seen only one movie add no links to the movie-to-movie network, the divergence of the weight $1/(k_i-1)$ at $k_i=1$ is not an obstacle.

Since between each pair of movies (α,β) we create multiple links, it is convenient to write their weights as a 5×5 matrix $\mathsf{W}_{\alpha\beta}$. Each rating can be represented by a column vector in 5-dimensional space: rating $v_{i\alpha}=1$ we represent as $\boldsymbol{v}_{i\alpha}=(1,0,0,0,0)^T$, rating $v_{i\alpha}=2$ as $\boldsymbol{v}_{i\alpha}=(0,1,0,0,0)^T$, and so forth. If the vote has not been given yet, we set $\boldsymbol{v}_{i\alpha}=(0,0,0,0,0)^T$. Then using the linking scheme from Fig. 2.4 and the user weights $1/(k_i-1)$ we write

$$\mathsf{W}_{\alpha\beta}=\sum_{i=1}^{M}\frac{\boldsymbol{v}_{i\alpha}\boldsymbol{v}_{i\beta}^T}{k_i-1}, \tag{2.1}$$

where we sum contributions from all users. In this way we convert the original data represented by a weighted bipartite graph into a weighted object-to-object network.

The non-normalized weights $\mathsf{W}_{\alpha\beta}$ form a symmetric matrix W with dimensions $5N\times 5N$. By the column normalization of W we obtain an unsymmetric matrix Ω. It describes a diffusion process on the underlying network with the outgoing weights from any node in the graph normalized to unity (see also a similar diffusion-like process in [68] and the PageRank algorithm[4]).

[3] Here one can recall the famous set of equations for PageRank $G(i)$ of webpage i. It has the form $G(i)=\alpha+(1-\alpha)\sum_{j\sim i}G(j)/k_j$, where the subscript j runs over all the webpages that contain a link to webpage i ($j\sim i$), for details see [40]. Here a similar scaling of the contributions by the inverse of the node degree arises. By a numerical solution of the set, one obtains values $G(i)$ which are essential for the Google search algorithm.

[4] Incidentally, PageRank algorithm normalizes the flux outgoing from a node in a similar way and thus it also represents diffusion or a random walk. If one chooses the row normalization instead, the resulting process is equivalent to heat conduction in the network.

Now we shall investigate the equation

$$\Omega \boldsymbol{h} = \boldsymbol{h}, \tag{2.2}$$

where $\boldsymbol{h}$ is a $5N$-dimensional vector (the first 5 elements correspond to movie 1, next 5 elements to movie 2, etc.). Denote $n_{\alpha s}$ ($\alpha = 1, \ldots, M$, $s = 1, \ldots, 5$) the number of times movie α has been rated with the rating s. Here we exclude the votes given by the users who have rated only one movie because these users do not contribute to Ω. It is easy to prove that the vector

$$\boldsymbol{h}^* = (n_{11}, \ldots, n_{15}, \ldots, n_{N1}, \ldots, n_{N5})^T \tag{2.3}$$

is a solution of Eq. (2.2). Moreover, the solution is unique up to multiplication by a constant and as we will see later, all vectors in the form $\lambda \boldsymbol{h}$, $\lambda \neq 0$, lead to identical predictions. Denote $L := 1 - \Omega$ the Laplace matrix, the forementioned uniqueness of $\boldsymbol{h}^*$ is equivalent to $\text{rank}(L) = 5N - 1$, which we prove in the following paragraph. It is worthwhile to emphasize that the unique solution $\boldsymbol{h}^*$ reproduces some features of the original input data, which strongly supports rationality and relevance of the construction of Ω.

Using elementary row/column operations one can shift all the rows/columns corresponding to the zero-rows/zero-columns of Ω to the bottom and right of L, leading to $\left(\begin{smallmatrix} L' & O \\ O & 1 \end{smallmatrix}\right)$, where O and 1 are the zero and the identity matrix. The dimension of 1 we label as D, the dimension of L' is then $5N - D$. The matrix L' has four properties: (i) All its diagonal elements are 1. (ii) All its non-diagonal elements lie in the range $[-1, 0]$. (iii) The sum of each column is zero. (iv) In each row, there is at least one non-diagonal nonzero element. One can prove that the rank of any matrix with these four properties is equal to its dimension minus one, $5N - D - 1$ in this case. Since $\text{rank}(1) = D$, together we have $\text{rank}(L) = \text{rank}(L') + \text{rank}(1) = 5N - 1$. Details of the proof will be shown in an extended paper.

The matrix Ω codes the connectivities between different ratings in the movie-to-movie network, and could yield to a recommendation for a particular user. Since the matrix represents only the aggregated information, in order to recommend for a particular user, we need to utilize opinions expressed by this user. We do so by imposing these ratings as fixed elements of $\boldsymbol{h}$ in Eq. (2.2). These fixed elements can be considered as a boundary condition of the given diffusion process; they influence our expectations on unexpressed ratings. In other words, large weights in Ω represent strong patterns in user ratings (*e.g.*, most of those who rated movie X with 5 gave 3 to movie Y) and diffusion of the ratings expressed by a particular user in the movie-to-movie network makes use of these patterns.

The discussion above leads us to the equation

$$\Omega_i \boldsymbol{h}_i = \boldsymbol{h}_i, \tag{2.4}$$

where $\Omega_i := \Omega$ for the rows corresponding to the movies unrated by user i and $\Omega_i := 1$ for the remaining rows. Such a definition keeps entries corresponding to the movies rated by user i preserved. The solution of Eq. (2.4) can be numerically obtained in a simple iterative way. We start with $\boldsymbol{h}_i^{(0)}$ where elements corresponding to the movies rated by user i are set according to these ratings and the remaining elements are set to zero. Then by the iteration equation $\boldsymbol{h}_i^{(n+1)} = \Omega_i \boldsymbol{h}_i^{(n)}$ we propagate already expressed opinions of user i over the network, eventually leading to the stationary solution $\boldsymbol{h}_i$. Intermediate results $\boldsymbol{h}_i^{(n)}$ contain information about the movies unrated by user i, which can give rise to a recommendation. We obtain the rating prediction as the standard weighted average. For example, if for a given movie in $\boldsymbol{h}_i$ we obtain the 5-tuple $(0.1, 0.2, 0.4, 0.3, 0.0)^T$, the rating prediction is $\hat{v} = 2.9$. Notice that if a user has rated no movies, we have to use a different method (for example the movie average introduced later) to make a prediction. This feature is common for recommender systems producing personalized predictions.

2.2.2 Avoiding the iterations

While simple, the iterative way to solve Eq. (2.4) has one important drawback: the iterations have to be made for every user separately. Consequently, the computational complexity of the algorithm is high. To get rid of this difficulty we rewrite Eq. (2.4) as $L\boldsymbol{h}_i = \boldsymbol{j}_i$, again $L = 1 - \Omega$. Here the external flux $\boldsymbol{j}_i$ is nonzero only for the elements representing the boundary condition of user i.

The solution $\boldsymbol{h}_i$ can be formally written in the form $\boldsymbol{h}_i = \mathsf{G}\boldsymbol{j}_i$. This resembles the well-known Green function approach: once G is known, $\boldsymbol{h}_i$ can be found by a simple matrix multiplication. While the source term $\boldsymbol{j}_i$ is not a priori known, we can get rid of it by reshuffling of the movies and grouping the boundary elements in $\boldsymbol{h}_i$. After this formal manipulation we obtain

$$\begin{pmatrix} \boldsymbol{h}_i^{\mathrm{B}} \\ \boldsymbol{h}_i^{\mathrm{F}} \end{pmatrix} = \begin{pmatrix} \mathsf{G}_{\mathrm{BB}} & \mathsf{G}_{\mathrm{BF}} \\ \mathsf{G}_{\mathrm{FB}} & \mathsf{G}_{\mathrm{FF}} \end{pmatrix} \begin{pmatrix} \boldsymbol{j}_i^{\mathrm{B}} \\ \boldsymbol{0} \end{pmatrix}, \tag{2.5}$$

where B stands for *boundary* and F for *free*. Now it follows that $\boldsymbol{h}_i^{\mathrm{B}} = \mathsf{G}_{\mathrm{BB}}\boldsymbol{j}_i^{\mathrm{B}}$ and $\boldsymbol{h}_i^{\mathrm{F}} = \mathsf{G}_{\mathrm{FB}}\boldsymbol{j}_i^{\mathrm{B}}$, leading us to the final result

$$\boldsymbol{h}_i^{\mathrm{F}} = \mathsf{G}_{\mathrm{FB}}\mathsf{G}_{\mathrm{BB}}^{-1}\boldsymbol{h}_i^{\mathrm{B}}. \tag{2.6}$$

Since most users have rated only a small part of all M movies, the dimension of G_{BB} is usually much smaller than that of G and thus the inversion $\mathsf{G}_{\mathrm{BB}}^{-1}$ is cheap.

The last missing point is that since the matrix L is singular (as mentioned above, $\mathrm{rank}(L) = 5N - 1$), the form of G can not be obtained by inverting L. Hence we use the *Moore-Penrose pseudoinverse* [69]

$$\mathsf{G} = L^{\dagger} = \lim_{k\to\infty} \left[1 + \Omega + \Omega^2 + \cdots + \Omega^k - k\boldsymbol{w}_{\mathrm{R}}\boldsymbol{w}_{\mathrm{L}}\right], \tag{2.7}$$

where $\boldsymbol{w}_{\mathrm{R}}$ and $\boldsymbol{w}_{\mathrm{L}}$ is the right and left eigenvector of Ω respectively, both corresponding to the eigenvalue 1. For practical purposes, the infinite summation in Eq. (2.7) can be truncated at a finite value k.

2.2.3 Personal polarization

Before the described method can be used in real life examples, there is one important technical problem. Each user has a different style of rating—some people tend to be very strict and on average give low marks, some people prefer to give either 1 or 5, some don't like to give low marks, and so forth. Thus, ratings cannot be grouped together in matrices $\mathsf{W}_{\alpha\beta}$ in the straightforward and naïve way we described before for they mean different things to different people.

To deal with this phenomenon, which we refer to as personal polarization, *unification* of ratings from different users is used before summing users' contributions in the object-to-object network. Consequently, before reporting resulting predictions to a user, the output of the algorithm has to be shifted back to the user's scale and *personalization* is needed.

To characterize the rating profile of user i we use the mean μ_i and the standard deviation σ_i of the votes given by him, and we compare these values with the mean m_i and the standard deviation s_i of the ratings given by all users. Notably, the quantities m_i and s_i take into account only the movies rated by user i—if a user has a low average rating because he has been rating only bad movies, there is no need to manipulate his ratings. To conform a user rating profile to the society rating profile we use the linear transformation

$$u_{i\alpha} = m_i + (v_{i\alpha} - \mu_i)\,\frac{s_i}{\sigma_i}. \tag{2.8}$$

Personalization of the predicted value is done by the inverse formula $v_{i\alpha} = \mu_i + (u_{i\alpha} - m_i)\sigma_i/s_i$. We can notice that while $v_{i\alpha}$ is an integer value, $u_{i\alpha}$ is a real number. Nevetheless, one can obtain its vector representation in the straightforward way: *e.g.*$u = 3.7$ is modelled by the vector $(0, 0, 0.3, 0.7, 0)^T$; the weighted mean corresponding to this vector is equal to the input value 3.7.

2.2.4 Benchmark methods

In correlation-based methods, rating correlations between users are quantified and utilized to obtain predictions. We present here one implementation of such a method, which serves as a benchmark for the proposed diffusion model. The correlation C_{ij} between users i and j is calculated with Pearson's formula

$$C_{ij} = \frac{\sum_\alpha^*(v_{i\alpha} - \mu_i)(v_{j\alpha} - \mu_j)}{\sqrt{\sum_\alpha^*(v_{i\alpha} - \mu_i)^2}\sqrt{\sum_\alpha^*(v_{j\alpha} - \mu_j)^2}}, \tag{2.9}$$

where we sum over all movies rated by both i and j (to remind this, there is a star added to the summation symbols); $C_{ij} := 0$ when users i and j have no movies in common. Due to the data sparsity, the number of user pairs with zero correlation can be high and the resulting prediction performance poor. To deal with this effect, in [70] it is suggested to replace the zero correlations by the society average of C_{ij}. In the numerical tests presented in this Letter the resulting improvement was small and thus we use Eq. (2.9) in its original form. Finally, the predictions are obtained using the formula

$$\hat{v}_{i\alpha} = \mu_i + \sum\nolimits'_j \frac{C_{ij}}{\sum'_k C_{ik}} (v_{j\alpha} - \mu_j). \tag{2.10}$$

Here we sum over the users who have rated movie α (prime symbols added to sums are used to indicate this), the term $\sum'_k C_{ik}$ serves as a normalization factor.

As a second benchmark method we use recommendation by the movie average (MA) where one has $\hat{v}_{i\alpha} = m_\alpha$, m_α is the average rating of movie α. This method is not personalized (for a given object, all users obtain the same prediction) and has an inferior performance. As it is very fast and easy to implement, it is still widely used. Notably, when unification-personalization scheme is employed together with MA, the predictions get personalized. As we will see later, in this way the prediction performance is increased considerably without a notable impact on the computation complexity.

2.2.5 Numerical results

To test the proposed method based on opinion diffusion (OD) we use the GroupLens project data, available at `www.grouplens.org`. The total number of users is $M = 943$, the total number of movies is $N = 1\,682$, and the ratings are integer values from 1 to 5. The number of given ratings is 100 000, corresponding to the voting matrix sparsity around 6%.

To test the described methods, randomly selected 10% of the available data is transfered to the probe file $\mathcal{P}$, and the remaining 90% is used as an input data for the recommendation. Then we make a prediction for all entries contained in the probe and measure the difference between the predicted value $\hat{v}_{i\alpha}$ and the actual value $v_{i\alpha}$. For an aggregate review of the prediction performance we use two common quantities: *root mean square error* (RMSE) and *mean absolute error* (MAE). They are defined as

$$\mathrm{MAE} = \frac{1}{n} \sum_{\mathcal{P}} |v_{i\alpha} - \hat{v}_{i\alpha}|, \tag{2.11a}$$

$$\mathrm{RMSE} = \left[\frac{1}{n} \sum_{\mathcal{P}} (v_{i\alpha} - \hat{v}_{i\alpha})^2 \right]^{1/2}, \tag{2.11b}$$

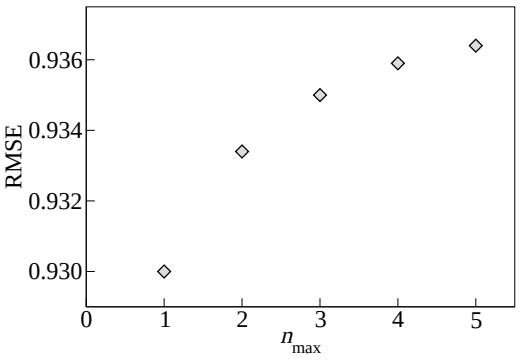


Figure 2.5: Prediction performance for the predictions $\hat{v}_{i\alpha}$ obtained by iterations of Eq. (2.4) using various numbers of iterations steps.

	no unification		with unification	
method	RMSE	MAE	RMSE	MAE
MA	1.10	0.87	0.98	0.77
CB	1.09	0.86	1.09	0.86
OD	1.00	0.80	0.93	0.73

Table 2.1: Comparison of the three recommendation methods: movie average (MA), correlation-based method (CB), and opinion diffusion (OD). Presented values are averages obtained using 10 different probes; standard deviations are approximately 0.01 in all investigated cases.

where the summations go over all user-movie pairs (i, α) included in the probe $\mathcal{P}$ and n is the number of these pairs in each probe dataset. To obtain a better statistics, the described procedure can be repeated many times with different selections of the probe data. We used ten repetitions which allowed us to compute not only the average performances (represented by RMSE and MAE) but also their standard deviations.

In contrast with the expectations, in Fig. 2.5 it can be seen that the prediction performance is getting worse by a small amount when more than one iteration of Eq. (2.4) is used to obtain the prediction. Probably this is due to the presence of overfitting—starting from the second iteration, our expectations are influenced not only by actually expressed ratings but also by our expectations about unexpressed ratings obtained in previous iteration steps. Nevertheless, as it will be shown later, the performance achieved by the first iteration is good and justifies validity of the proposed model. In the following paragraphs we use only one iteration to obtain the predictions. Consequently, the Green function method introduced above is not necessary—we decided to expose it in this paper because it can be useful with other datasets.

In Tab. 2.1 we compare the prediction accuracy for the movie-average method (MA), the correlation-based method (CB), and for the opinion diffusion (OD). To measure the prediction performances we use both RMSE and MAE as defined above. All three methods are tested both with and without employing the unification-personalization scheme. In accordance with expectations, for MA and OD the performances with unification included are better than without it; for the simplest tested method, MA, the difference is particularly remarkable. By contrast, CB is little sensitive to the unification procedure and when we drop the multiplication by σ_i/s_i from the unification-personalization process given by Eq. (2.8), the difference disappears completely (which can be also confirmed analytically). According to the prediction performances shown in Tab. 2.1 we can conclude that the diffusion method outperforms the other two clearly in all tested cases (RMSE/MAE, with/without unification). When computation complexity is taken into account, it can be shown that if $M > N$, the proposed method is more effective than correlation-based methods (but, of course, less effective than using the movie average).

2.2.6 Conclusion

We have proposed a novel recommendation method based on diffusion of opinions expressed by a user over the object-to-object network. Since the rating polarization effect is present, we have suggested the unification-personalization approach as an additional layer of the recommender system. To allow a computation reduction with some datasets, Green function method has been introduced. The proposed method has been compared with two standard recommendation algorithms and it has achieved considerably better results. Notably, it is executable even for the large dataset (17 770 movies, 480 189 users) released by Netflix (a DVD rental company, see `www.netflixprize.com`). In addition, our model is tune-free in essence—it does not require extensive testing and optimization to produce a high-quality output. This is a good news for practitioners.

2.3 Comparison of recommendation methods

Courage is going from failure to failure without losing enthusiasm.

Winston Churchill

In this section we overview several standard recommendation methods and compare their accuracies on various datasets. To assess the prediction accuracy, a small part of the input data is removed and constitutes the *probe* dataset. The rest, *training* dataset, is then used as input for the tested methods. Compar-

ing the predicted values with the corresponding probe entries, the root mean square error (RMSE) can be calculated as given in (2.11ab). The usual way to create the probe is by random selecting 10% of the input data. However, users who have given many ratings then comprise a large part of the probe and the results may be biased towards them. For this reason we limit number of probe entries per user to ten. This choice leads to higher resulting RMSEs (compared with the random 10%-selection) but it is more user-focused and hence more representative.

For our tests we use various datasets, all of them containing ratings in the integer scale $1, 2, 3, 4, 5$. Dataset A is the smaller one of the two datasets which have been prepared by the GroupLens project.[5] Datasets B and C are based on the original Netflix dataset. Dataset B contains randomly selected 5 922 movies and 10 000 users who have rated at least 10 movies, by the random selection we aim to capture properties of the large original data by a smaller subset. Dataset C contains 3 000 movies which have $[1\,400; 5\,400]$ ratings and 3 000 users who have given $[1\,250; 1\,500]$ ratings, this choice selectively picks a denser part (not the densest!) of the original data. Finally dataset D is based on the data prepared by the Eigentaste project, it contains ratings of 3 000 users for 100 jokes.[6] Since the original ratings use the real scale $[-10; 10]$, they have been transformed by the simple linear mapping $v \mapsto 3 + v/4$ and coarse-grained by rounding to integer values.

Basic statistical properties of the datasets are summarized in Tab. 2.2. The code given in the first column is used to refer to the dataset. The following three columns give the number of users, objects, and ratings respectively. The data sparsity is the ratio of the number of ratings to the maximum possible number of ratings which is users $\times$ movies. In the last two columns are the minimal number of ratings by a user (user_min) and the minimal number of ratings of a movie (movie_min). As can be seen, the joke dataset D is much denser then the other three datasets, with all objects having a large number of ratings. As we will see later, higher density of the input data does not automaticaly provide better prediction precision, the inner structure of the data can reverse the results (for more details see [70]).

2.3.1 Standard recommendation methods

Before proceeding to numerical results, we describe the tested methods in detail. Opinion diffusion is omitted here because it is implemented exactly as described in Sec. 2.2 (with one diffusion step).

[5] For more information see `www.grouplens.org`.

[6] On the webpage `eigentaste.berkeley.edu` one can both appreciate joke recommendation and obtain more information about their research.

code	users	objects	ratings	sparsity	user_min	object_min
A	943	1 682	100 000	6.30%	20	1
B	10 000	5 922	824 802	1.37%	10	1
C	3 000	3 000	567 456	6.31%	45	23
D	3 000	100	243 380	81.13%	36	1 160

Table 2.2: Basic statistical properties of the used datasets.

Object and user averages

Denoting the average rating given to object α as μ_α, a simple prediction for user i can be obtained as

$$\hat{v}_{i\alpha} = \mu_\alpha. \tag{2.12}$$

As described in Sec. 2.2, this plain method can be improved by the unification-repersonalization process. That means, denoting the average rating given by user i as μ_i, the standard deviation of ratings given by user i as σ_i, the average rating of all users on the subset of objects rated by user i as m_i, and the standard deviation of all users on this subset as s_i, the ratings can be unified by Eq. (2.8). Then for each object, the unified average rating μ'_α can be computed. By transforming back to the personal scale of user i we obtain the prediction

$$\hat{v}_{i\alpha} = \mu_i + \left(\mu'_\alpha - m_i\right) \frac{\sigma_i}{s_i}. \tag{2.13}$$

This variation of the recommendation by object average we indicate with an asterisk (*).

Similarly simple is the prediction by μ_i, the average rating of user i,

$$\hat{v}_{i\alpha} = \mu_i. \tag{2.14}$$

Here the application of the unification-personalization process is not justified and hence in our numerical tests we consider only this basic method.

User-user and object-object correlations

Similarities of users can be quantified by computing correlations between their ratings, the standard choice being the Pearson correlation coefficient

$$c_{ij} = \frac{1}{|\mathcal{C}_{ij}|} \sum_{\alpha \in \mathcal{C}_{ij}} \frac{(v_{i\alpha} - \mu_i)(v_{j\alpha} - \mu_j)}{\sigma_i \sigma_j} \tag{2.15}$$

where $\mathcal{C}_{ij}$ is the set of movies rated by both i and j. Consequently, the prediction $\hat{v}_{i\alpha}$ can be obtained as a weighted average

$$\hat{v}_{i\alpha} = \frac{\sum'_j \lambda_{ij} v_{j\alpha}}{\sum'_j \lambda_{ij}}. \tag{2.16}$$

where we sum only over the users who have rated object α and the weights λ_{ij} depend on the user-user-correlations c_{ij}. While in the literature, several possible choices of weights are discussed [59, 70, 71], in our numerical tests the best performance was achieved by the simple formula $\lambda_{ij} = c_{ij}\theta(c_{ij})\sqrt{|\mathcal{C}_{ij}|}$. That is, we neglect negative correlations and multiply the weight by the square root of the number of movies rated by both i and j (c_{ij} obtained with a larger overlap has a larger statistical weight).

Correlations of object ratings can be employed in a very similar way. First we calculate the correlations

$$c_{\alpha\beta} = \frac{1}{\mathcal{C}_{\alpha\beta}} \sum_{i \in \mathcal{C}_{\alpha\beta}} \frac{(v_{i\alpha} - \mu_i)(v_{i\beta} - \mu_i)}{\sigma_\alpha \sigma_\beta} \tag{2.17}$$

and then the weighted averages

$$\hat{v}_{i\alpha} = \frac{\sum'_\beta \lambda_{\alpha\beta} v_{i\beta}}{\sum'_\beta \lambda_{\alpha\beta}} \tag{2.18}$$

where we sum only over the movies β that have been rated by user i. In our tests, good performance was obtained by $\lambda_{\alpha\beta} = c_{\alpha\beta}\theta(c_{\alpha\beta})\sqrt{|\mathcal{C}_{\alpha\beta}|}$.

One could try to improve the correlation-based methods described above by unification-personalization process. According to our numerical tests such efforts are fruitless, yielding accuracy increase (measured by RMSE) less than one percent for both user-user and object-object correlations.

Matrix factorization

There is a large class of techniques based on singular value decomposition. The implementation described here is based on the algorithm used in the Netflix-prize (see `www.netflixprize.com`) by team Gravity which hold the leading position for several months [59].

Labeling the voting matrix as V, we try to approximate it by the product of two matrices

$$\mathsf{V} \approx \mathsf{U}\mathsf{O} \tag{2.19}$$

where U is a $M \times K$ matrix, O is a $K \times N$ matrix, M is the number of users, N is the number of objects, and K is an integer parameter of the model. Values u_{ik} ($k = 1, \ldots, K$) can be considered as K tastes of user i, values $o_{k\alpha}$ ($k = 1, \ldots, K$) as K features of object α. Notice that it is implicitely assumed that rating of user i for object α is obtained as the inner product $\sum_{k=1}^{K} u_{ik} o_{k\alpha}$ of tastes and features. Consequently, K can be described as the number of dimensions which is used to characterize tastes and features.

For a given U and O we can measure the differences between UO and actual ratings $\mathcal{R}$ contained in the voting matrix as

$$e_{i\alpha} = v_{i\alpha} - \sum_{k=1}^{K} u_{ik} o_{k\alpha}, \tag{2.20}$$

$$\mathrm{TE} = \sum_{(i,\alpha)\in\mathcal{R}} e_{i\alpha}^2. \tag{2.21}$$

Here $e_{i\alpha}$ is the error for one particular rating and TE is the total error obtained by summing over all ratings contained in the voting matrix V. Square root of this value is often called *training RMSE* because it measures how well we approximate the "training" data stored in the voting matrix.

To find the matrices U, O that describe the actual ratings well, one should try to minimize the total error TE. This can be achieved by applying the gradient descent method. That means, for any $(i,\alpha) \in \mathcal{R}$ we compute the gradients of $e_{i\alpha}^2$

$$\frac{\partial e_{i\alpha}^2}{\partial u_{ik}} = -2 e_{i\alpha} o_{k\alpha}, \quad \frac{\partial e_{i\alpha}^2}{\partial o_{k\alpha}} = -2 e_{i\alpha} u_{ik}$$

and update the corresponding elements in the opposite direction of the gradients

$$u_{ik} \to u_{ik} + \eta\,(2 e_{i\alpha} o_{k\alpha}), \quad o_{k\alpha} \to o_{k\alpha} + \eta\,(2 e_{i\alpha} u_{ik}). \tag{2.22}$$

The small parameter η is usually called learning rate, it controls how fast we modify U and O. To prevent large values of u_{ik} and $o_{k\alpha}$, in [59] the damping parameter λ is introduced to the iterations as

$$u_{ik} \to u_{ik} + \eta\,(2 e_{i\alpha} o_{k\alpha} - \lambda u_{ik}), \quad o_{k\alpha} \to o_{k\alpha} + \eta\,(2 e_{i\alpha} u_{ik} - \lambda o_{k\alpha}). \tag{2.23}$$

As we will check by numerical simulations, $\lambda > 0$ indeed yields better accuracy and the improvement is substantial, more than 6% (see Tab. 2.3). However, after introduction of λ, interpretation of the method is less clear because stationary points of Eq. (2.23) do not relate to $e_{i\alpha}$ anymore.

To conclude, we summarize the matrix factorization method in steps.

1. Initialize U and O randomly, set small η and λ.
2. While RMSE obtained using the probe dataset decreases, repeat:
   a) For each $(i,\alpha) \in \mathcal{R}$ compute $e_{i\alpha}$ and update matrices U, O according to Eq. (2.23).
   b) Calculate RMSE on the probe dataset.

While the idea behind the method is intuitively appealing, it has an important drawback—to optimize the prediction performance, parameters K, η, λ have to be tuned and the consequent optimization procedure involves lots of numerical

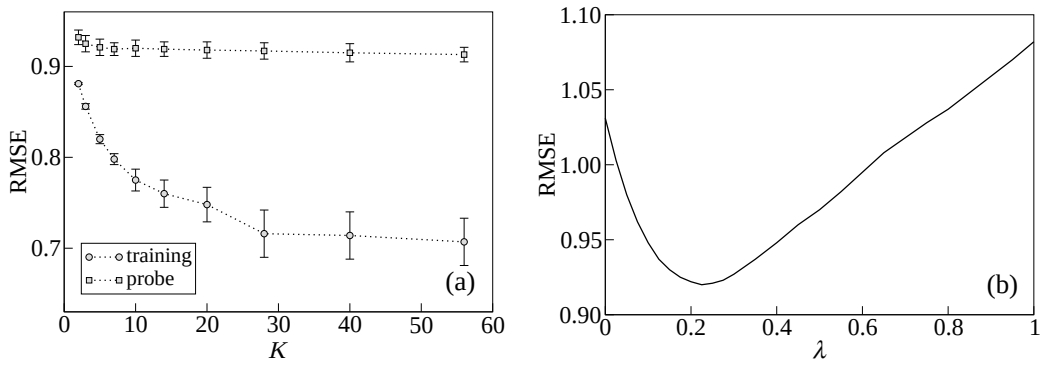


Figure 2.6: (a) Prediction accuracy vs the number of hidden dimensions K for $\eta = 0.001$ and $\lambda = 0.2$, both training and probe RMSEs are shown. (b) Probe RMSE vs the optimization parameter λ. In both figures, dataset A was used and the results are averaged over 10 different probes.

tests. Moreover, the optimal values are not universal and for a different dataset they must be optimized again. To better understand how the method behaves, we examine how the prediction accuracy (measured by RMSE) depends on the parameters K, η, λ for the dataset A. In Tab. 2.3 we show probe RMSE for various η and λ. As can be seen, there is an optimal value for λ which is in our case $\lambda \approx 0.2$. By contrast, accuracy decreases with η and thus there is no optimal value. Consequently, the chosen η is a trade-off between accuracy and computation speed (the number of iterations needed to reach the minimal RMSE is inversely proportional to η).

In Fig. 2.6a we show both training and probe RMSE as functions of K for fixed $\eta = 0.001$ and $\lambda = 0.2$. As can be seen, while increasing K allows us to better reproduce the known ratings (*i.e.*, it lowers the training RMSE), it does not improve the prediction accuracy substantially (*i.e.*, the probe RMSE does not decrease much).[7] Since computation speed is inversely proportional to K, this is not a bad news—even a small value of K allows for almost optimal performance. In Fig. 2.6b we show how the prediction performance depends on the choice of λ. As can be seen, λ is an important parameter which needs to be set very carefully. Since $\lambda = 0$ does not produce plausible results, the iterative optimization of Eq. (2.21) cannot be replaced by some of the stochastic methods for global optimization in multidimensional spaces (simulated annealing, genetic algorithms, etc.).

[7] For some datasets, the probe RMSE even grows when K is larger than a certain critical value (see *e.g.* [59]). This is usually referred to as *overfitting*.

$\eta \diagdown \lambda$	0.0	0.1	0.2	0.3	0.4
0.0001	0.984	0.947	0.920	0.926	0.947
0.0003	0.984	0.947	0.920	0.926	0.947
0.0010	0.984	0.947	0.921	0.926	0.949
0.0030	0.988	0.950	0.924	0.931	0.955
0.0100	1.033	0.974	0.947	0.962	0.990

Table 2.3: Prediction accuracy for various values of η and λ, $K = 20$ (dataset A, results for one probe only).

method $\diagdown$ dataset	A	B	C	D
object average	1.02	1.03	1.00	1.27
object average*	0.96	0.96	0.87	1.11
user average	1.04	1.02	0.97	1.16
user correlations	1.06	1.07	0.98	1.17
object correlations	1.00	0.98	0.91	1.13
matrix factorization	0.92	0.91	0.82	1.04
opinion diffusion	0.98	0.97	0.91	1.19
opinion diffusion*	0.94	0.94	0.85	1.09

Table 2.4: Accuracies (measured by RMSE) of the tested recommendation methods. Values are averages over to different probes, standard deviations are less than 0.01 in all cases.

2.3.2 Results

Accuracies of the tested recommendation methods, as measured by RMSE, are shown in Tab. 2.4. Values are obtained by averaging over ten different probes, standard deviations are less than 0.01 in all cases (it decreases with the probe size, as expected). We remind that asterisk-marked methods use the unification-personalization process.

Finally, values shown in Tab. 2.4 allow us to draw several conclusions. First, correlation-based methods, while intuitively appealing, yield accuracies which are comparable with accuracies of plain averages (object and user average). Since calculation of correlations is much more demanding then calculation of averages, correlation-based methods are clearly least advisable among the tested methods. Second, when object average is supplemented by the unification-personalization process, it performs remarkably well and there are only two methods (matrix factorization and opinion diffusion*) which are superior to it. In addition, the computation complexity and memory requirements of object average* is much lower than for the two superior methods and hence it is a very good candidate for "cheap" recommendation method. Third, matrix fac-

torization was the most successful method, it outperformed opinion diffusion* on all four datasets by 2%–5%. This is not surprising because opinion diffusion* is tune-free in essence and hence it can lag behind methods with several tunable parameters.[8] This gives us a suggestion for future research: to amend opinion diffusion with a small set of parameters which, when tuned appropriately, could increase the prediction accuracy.

2.4 Application of diffusion to bipartite networks

While ratings of objects by users are naturally represented by a weighted bipartite network, many other cases result in unweighted bipartite networks [72–80]. This motivates us to adapt the diffusion method to unweighted bipartite networks.

2.4.1 Projections of bipartite networks

Nodes of each bipartite network can be divided into two groups, let's label these groups as X and Y, such that links are only between nodes that do not belong to the same group. To simplify such a network, *projections* on X-network (X-projection where only the nodes from group X are present) or Y-network (Y-projection where only the nodes from group Y are present) are used. For example, in the simplest X-projection two nodes are connected when in the original network they have at least one common neighbouring Y-node. Notice that such a projection results in an unweighted network; for other studies of unweighted projections of bipartite networks see [74, 75, 81–83]. Although some topological properties can be qualitatively obtained from this unweighted version, the loss of information is obvious: if two listeners have collected more than 100 music groups each and only one music group is selected by both listeners, these two listeners probably have quite different music taste. By contrast, if nearly 100 music groups belong to the overlap, the two listeners are likely to have very similar habits. However, in the unweighted listener-projection, this two cases have exactly the same graph representation.

Weighted projections reduce the loss of information. A straightforward way is to weight each edge by the number of of common neighbours in the original network [84–86]. However, this method is also quantitatively biased. For example, in the empirical study of scientific collaboration networks [87] it is suggested that the impact of one additional collaboration paper should depend on the original weight between the two scientists. For example, one more co-authorized paper for the two authors having only co-authorized one paper be-

[8] When taken to extreme, in addition to K, λ, η, all elements of the matrices U, O can also be considered as tunable parameters.

fore should have higher impact than for the two authors having already co-authorized 100 papers. This saturation effect can be taken into account by introducing a hyperbolic tangent function onto the simple count of collaborated times [87]. Another suggestion is that two scientists whose names appear on a paper together with many other co-authors know one another less well on average than two who were the sole authors of a paper [75]. Labeling the number of co-authors as n_C, the factor $1/(n_C - 1)$ can weaken the contribution of huge collaborations [88, 89].

However, the solutions above are chosen *ad hoc*, *e.g.*instead of the hyperbolic tangent, many other candidates can be used. Weighting by the factor $1/(n_C - 1)$ excludes the papers with only one author which is often the case in some sciences (for example, more than half of the publications in *Mathematical Reviews* have only one author [81]). In addition, for simplicity, the weighted adjacency matrix $\{w_{ij}\}$ is usually set to be symmetrical, that is, $w_{ij} = w_{ji}$. Obviously, in real scientific collaboration networks, different authors may assign different weights to the same co-authorized paper—it is probably the case that the author having less publications may give a higher weight and vice versa. Hence, it is natural to relax this constraint and search for an asymmetrical weighting method.

In the following we propose a weighting method with asymmetrical weights and allowed self-connection (*i.e.*, $w_{ii} > 0$). This method can be directly applied as a personal recommendation algorithm, which performs remarkably better than the widely used *global ranking method* (GRM) and *collaborative filtering* (CF).

2.4.2 Method

In the prospective X-projection, the weight w_{ij} can be considered as the importance of node i in j's sense and in general it differs from w_{ji}. For example, in the book-projection of a customer-book network, the weight w_{ij} between two books i and j contributes to the strength of book i recommendation to a customer provided he has brought book j. We assume that before the projection is made, a certain amount of a resource (*e.g.*, social influence in a recommendation system) is associated with each X-node. Consequently, the weight w_{ij} represents the proportion of the resource j would like to distribute to i.

A suitable form of w_{ij} can be obtained by studying the original bipartite network. Since the network is unweighted, the unbiased allocation of the initial resource of each X-node is equally among all its neighboring Y-nodes. Subsequently, the resource collected on Y-nodes are equally redistributed back to their neighboring X-nodes. An illustration of this resource-allocation process for a simple bipartite network is shown in Fig. 2.7. There, the three X-nodes (up) are initially assigned weights x, y and z. In the first step, from Y-nodes (bottom). Merging the two steps in one, the final resources (x', y', z') located on

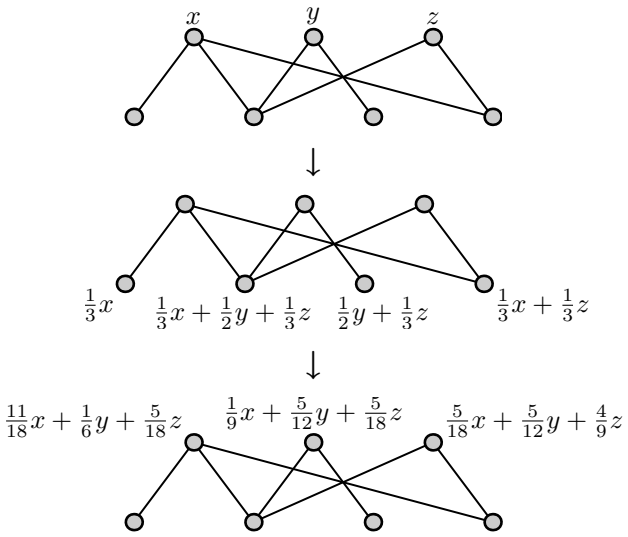


Figure 2.7: Illustration of the resource-allocation process in a simple bipartite network. The assigned resources first flow from X-nodes to Y-nodes and then return back to X-nodes.

X-nodes can be obtained as

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 11/18 & 1/6 & 5/18 \\ 1/9 & 5/12 & 5/12 \\ 5/18 & 5/12 & 4/9 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{2.24}$$

Notice that the induced 3×3 matrix is column normalized, the element in the ith row and jth column represents the fraction of resource the jth X-node transferred to the ith X-node. To conclude, the described resource-allocation process yields a weighted adjacency matrix which captures relations between X-nodes. In other words, we obtained a weighted X-projection of the original bipartite network.

To generalize the concept explained above, let's consider a bipartite network $G(X, Y, E)$, where X and Y are two groups of nodes and E is the set of edges. The nodes in X are denoted by $i = 1, \ldots, n$, the nodes in Y by $\alpha = 1, \ldots, m$. The initial resource located on the i-th X-node is $x_i \geq 0$. After the first step, all the resource in X flows to Y and the resources y_α on Y-nodes read

$$y_\alpha = \sum_{i=1}^{n} \frac{a_{i\alpha} x_i}{k_i} \tag{2.25}$$

where k_i is the degree of node i and $a_{i\alpha}$ is the adjacency matrix

$$a_{i\alpha} = \begin{cases} 1 & (i, \alpha) \in E, \\ 0 & \text{otherwise.} \end{cases} \tag{2.26}$$

In the next step, all resources flow back to X-nodes and the final resources x'_i read

$$x'_i = \sum_{\alpha=1}^{m} a_{i\alpha} y_\alpha / k_\alpha = \sum_{\alpha=1}^{m} \frac{a_{i\alpha}}{k_\alpha} \sum_{j=1}^{n} \frac{a_{j\alpha} x_j}{k_j}. \tag{2.27}$$

This can be rewritten as

$$x'_i = \sum_{j=1}^{n} w_{ij} x_j, \quad w_{ij} = \frac{1}{k_j} \sum_{\alpha=1}^{m} \frac{a_{i\alpha} a_{j\alpha}}{k_\alpha}. \tag{2.28}$$

In effect, w_{ij} contains contributions from all 2-step paths between i and j. The matrix $\mathsf{W} = \{w_{ij}\}_{n\times n}$ represents the obtained weighted X-projection of the original network. Using matrix formalism, the process can be described by the linear equation $x' = \mathsf{W}x$.

It is worthwile to discuss some features of the proposed weighting method. For convenience, we take the scientific collaboration network as an example but our statements are valid also for other networks. First, W is not symmetric as

$$\frac{w_{ij}}{k_j} = \frac{w_{ji}}{k_i}. \tag{2.29}$$

This agrees with our daily experience: the weight of a single collaboration paper is relatively small if the scientist has already published many papers (*i.e.*, has a large degree) and vice versa. Second, the diagonal elements of W are nonzero and thus the information contained in links to Y-nodes with degree one is not lost. Actually, the diagonal element is the maximal element in each column. Only when all neighbours of node i are also neighbours of j, $w_{ii} = w_{ji}$ (it can occur that a student coauthors all papers with her supervisor). Consequently, the ratio $w_{ji}/w_{ii} \leq 1$ can be considered as research independence of i from j: the smaller the ratio, the more independent the researcher is and vice versa. Eventually, the global independence of i can be measured as

$$I_i = \sum_{j=1}^{n} (w_{ji}/w_{ii})^2. \tag{2.30}$$

Generally speaking, an author who often publishes papers alone or publishes many papers with different coauthors is more independent. To obtain a better understanding of this quantity, an extensive study of empirical data would be necessary. Instead, we return to our original aim, to recommender systems.

2.4.3 Personal recommendation

Let's label objects in the object-set by $\alpha = 1, \ldots, n$ and users in the user-set by $i = 1, \ldots, m$. If users are only allowed to collect objects (*i.e.*, they do not rate

them), their previous opinions can be stored in an $n \times m$ adjacency matrix $\{a_{\alpha i}\}$ where $a_{\alpha i} = 1$ if user i has already collected object α, and $a_{\alpha i} = 0$ otherwise. We assume that the collected objects are appreciated by the users who collected them. Consequently, the goal of a recommendation method is for each user to pick up those uncollected objects which are likely to by appreciated by the given user.

The degree of object α is $k_\alpha = \sum_{i=1}^{m} a_{\alpha i}$. The simplest available recommendation method, *global ranking method* (GRM), sorts all objects in the descending order of degree and recommended are the objects with highest degrees. Although such a recommendation is not personalized which in turns leads to poor performance of this method (see numerical comparison in the next section), the method is widely used because it is simple and spares computational resources. For example, the well-known *Yahoo Top 100 MTVs*, *Amazon List of Top Sellers*, as well as lists of most downloaded articles in scientific journals, can all be considered as results of GRM.

Another widespread recommendation algorithm, *collaborative filtering* (CF), is based on user-user similarities [48, 90]. The similarity of users i and j can be introduced in the Pearson-like form

$$s_{ij} = \frac{\sum_{\alpha=1}^{n} a_{\alpha i} a_{\alpha j}}{\min\{k_i, k_j\}} \tag{2.31}$$

where $k_i = \sum_{\alpha=1}^{n} a_{\alpha i}$ is the degree of user i and $s_{ij} = s_{ji}$. For any user-object pair (i, α), if user i has not collected object α yet (*i.e.*, $a_{\alpha i} = 0$), the score $v_{i\alpha}$ assigned by CF (quantifying how likely α will be appreciated by i) is

$$v_{i\alpha} = \frac{\sum_{j=1, j\neq i}^{m} s_{ij} a_{\alpha j}}{\sum_{j=1, j\neq i}^{m} s_{ij}}. \tag{2.32}$$

There are two factors contributing to a high value of $v_{i\alpha}$. First, if the degree of object α is high, there are many nonzero terms in the numerator of Eq. (2.32). Second, if object α is frequently collected by users similar to user i, the resulting value $v_{i\alpha}$ is large. The former effect accounts with the global information, the latter effect provides personalization. Finally, for a given user i, uncollected objects with highest values $v_{i\alpha}$ are recommended.

We propose a recommendation algorithm, which is a direct application of the weighting method presented above. The layout is simple. First we compress the bipartite user-object network by object-projection; the resulting weighted network we label as G. Then for a given user i we put some resources on the objects already collected. For simplicity we set the initial resource located on each node of G as

$$o_\alpha = a_{\alpha i}. \tag{2.33}$$

Thus, if object α has been collected by i then its initial resource is one, otherwise it is zero. We emphasize that the initial configuration is different for different

users and, similarly to the model presented in Sec. 2.2, it can be understood as the initial condition in a diffusion process. According to the resource-allocation process discussed above, the final resources are

$$o'_\alpha = \sum_{\beta=1}^{n} w_{\alpha\beta} o_\beta = \sum_{\beta=1}^{n} w_{\alpha\beta} a_{\beta i}. \tag{2.34}$$

Finally, the objects uncollected by the given user i are sorted according to o'_α and objects with highest values are recommended. We call this method *network-based inference* (NBI), since it is based on the weighted network G. Note that, the calculation of Eq. (2.34) have to be repeated m times (for each user).

2.4.4 Numerical results

For numerical tests of the described algorithms we use the dataset provided by *MovieLens* project (see `http://www.grouplens.org`). The dataset contains ratings of 1 682 movies (objects) by 943 users. Since the original data use the integer rating scale $1, 2, 3, 4, 5$, a coarse-graining procedure is necessary to obtain an unweighted bipartite network. As in [71], we assume that a movie has been collected by a user if the given rating is at least 3. If there is no rating or the rating is 1 or 2, no link is drawn. Out of the original 10^5 ratings, $85\,250$ ratings are at least 3 and thus the resulting sparsity of the adjacency matrix is only 5%. To test the recommendation algorithms, the data set (*i.e.*, 85 250 edges) is randomly divided into two parts: The training set contains 90% of the data, and the remaining 10% of data constitutes the probe. The training set is treated as known information, while no information from the probe set is allowed to be used for prediction.

All three algorithms (GRM, CF, and NBI) provide a sorted list of recommended objects for each user. For a user i, if a recommended object is in the probe set, we quantify its position in the list. For example, if user i has 1 500 uncollected movies, his recommendation list has the length 1 500. Consequently, if a recommended object α is on 30th position from the top, we say that it's a top 30/1 500 object and denote it by the relative rank $r_{i\alpha} = 20/1500 = 0.02$. Since the probe entries are actually collected by users, a good algorithm is expected to place probe entries on top of recommended lists, leading to small average rank r. Numeric results, averaged over all entries in the probe (and over different probes), are 0.139, 0.120 and 0.106 by GRM, CF, and NBI respectively. Since NBI yields the smallest value, we can say that among the three tested methods, NBI performs best. Notably, the difference between NBI and GRM is approximately the same as the difference between sophisticated CF and plain GRM.

To make our comparison more relevant to real-life recommendation systems, we introduce a measure of algorithmic accuracy that depends on the length of recommendation list. For a particular user, the recommendation list with the

length	GRM	CF	NBI
10	10.3%	14.1%	16.2%
20	16.9%	21.6%	24.8%
50	31.1%	37.0%	41.2%
100	45.2%	51.0%	55.9%

Table 2.5: The hitting rates for some typical lengths of recommendation list.

length L contains L highest recommended movies. For each movie α which is included in the probe dataset we count one hit. The ratio of hit entries to the probe size we call *hitting rate* (*i.e.*, it quantifies the fraction of the probe dataset discovered by recommendations). For a given L, the algorithm with a higher hitting rate is better and vice versa. Obviously, the hitting rate grows with L with the upper bound 1 for a sufficiently large L. The resulting hitting rates for all three methods and various lengths L are shown in Tab. 2.5. Again, NBI performs significantly better than GRM and CF.

2.4.5 Conclusion

Based on a particular projection of bipartite networks we proposed a recommendation method. We measured its recommendation performance by two different quantities and the obtained results we compared with results of two other standard methods. According to both quantities, the proposed method performs best. We emphasize that this model is only a first proposal, various modifications can be explored to increase recommendation performance. For example, the generalization of the initial condition $o_\alpha = a_{\alpha i}$ to the form $o_\alpha = a_{\alpha i} k_\alpha^\gamma$ is investigated (the exponent γ is a free tunable parameter) in [63]. In [91], the simple rule $o' = \mathsf{W}o$ is replaced by $o' = (\mathsf{W} + \lambda_2 \mathsf{W}^2 + \lambda_3 \mathsf{W}^3)o$ which effectively introduces repetitions of the resource-allocation process; optimization of $\lambda_{2,3}$ yields $\lambda_2 < 0$ and $\lambda_3 = 0$.

If we denote $\langle k_u \rangle$ and $\langle k_o \rangle$ the average degree of users and objects in the bipartite network, the computational complexity of CF is $\mathbb{O}(m^2 \langle k_u \rangle + mn \langle k_o \rangle)$ where the first term accounts for the calculation of similarity between users by Eq. (2.31) and the second term accounts for the calculation of the predicted score by Eq. (2.32). Since $n\langle k_o \rangle = m\langle k_u \rangle$ (both sides are equal to the total number of edges in the network), we are left with $\mathbb{O}(m^2 \langle k_u \rangle)$. The computational complexity of NBI is $\mathbb{O}(m\langle k_u^2 \rangle + mn\langle k_u \rangle)$ with two terms accounting for the calculation of the weighted matrix $w_{\alpha\beta}$ and the final resources o'_α respectively. Here $\langle k_u^2 \rangle$ is the second moment of the users' degree distribution in the bipartite network. Clearly, $\langle k_u^2 \rangle < n\langle k_u \rangle$ and thus the complexity above can be simplified to $\mathbb{O}(mn\langle k_u \rangle)$. Consequently, in systems where the number of users is much

larger than the number of objects, NBI runs much faster than CF. In addition, NBI requires n^2 memory to store the weighted matrix $\{w_{ij}\}$, while CF requires m^2 memory to store the similarity matrix $\{s_{ij}\}$. Hence, NBI is able to beat CF in all the three criterions of recommendation algorithm: *accuracy*, *time*, and *space*. However, in some recommendation systems, as in bookmark sharing websites, the number of objects (*e.g.*, webpages) is much larger than the number of users, thus CF may be more applicable.

3 Probabilistic approaches to market modeling

Human beings, viewed as behaving systems, are quite simple. The apparent complexity of our behavior over time is largely a reflection of the complexity of the environment in which we find ourselves.

Herbert Simon

Vendors and buyers interacting in a market present another example of a complex system. The complexity stems from various sources: many products and product variants which differ in features and prices are simultaneously available, buyers are diverse, and the decision process of every singly buyer is complex itself [92, 93]. The standard economics approach to modeling buyer decisions in a market is built on maximization of the expected utility by each party involved in the system (be it a buyer or a vendor). Since complexity of the problem is often enormous, there are serious concerns whether real customers can act in a way described by this game-theoretic models [94]. However, it is not our purpose here to criticize the use of the utility theory—an interested reader can find deeper discussions works devoted to this topic in *e.g.* [93–95].

When a buyer forms an opinion about a product, quality and features of the product jointly influence the outcome of buyer's consideration (others factors, *e.g.*price of the product, play also a role). We go in the direction of models of discrete choice[1] [96–98] and assume that customers' decisions, while influenced by perceived properties of the products, are probabilistic in nature. Using a probabilistic consumer choice framework makes it possible to avoid utility functions and hence our model can be understood as an alternative to the usual utility-function approach.

We developed two novel market models where buyers' decisions are probabilistic. The first model focuses on product quality (Sec. 3.1), the second on matching products' features with buyers' preferences (Sec. 3.2). While in reality neither of the two influences can be neglected, the "divide & conquer" tactics allow us to show that our simple assumptions yield rich behavior. The resulting complex complex systems with one vendor, several product variants, and

[1]The label "models of probabilistic choice" is also used.

many heterogeneous buyers, are investigated by both analytical and numerical techniques.

The discussed models were constructed to investigate effects of correlations and limited information in complex systems. Although they are presented in the context of economics, they provide general concepts and tools which can be used to investigate many other problems where involved parties have aligned, partially aligned, or antagonistic interests. We plan to use the established methodology for our future research.

Before we proceed, it's useful to make a clear distinction between quality and preferences. Let's imagine a large group of customers who are all asked to review a given product. While the assessments naturally differ, their common part can be attributed to the product's quality and their variable part (which differs from one customer to another) can be attributed to matching between customers' tastes and product's attributes. In theoretical models, this matching is often represented by a scalar product of the vector of customer's tastes and the vector of product's attributes, see *e.g.* [57, 99].

3.1 Selection by quality

Science is simply common sense at its best.

Thomas Huxley

The interests of vendors and customers seem antagonistic *a priori*, the former aiming at decreasing quality and increasing price, whereas the latter wishing exactly the opposite. The situation is fortunately more complex, the interests of both sides being sometimes compatible. Intuitively, a vendor may sell more items by increasing their perceivable quality, making everybody happier. But the situation is more subtle because of asymmetric information: the vendor knows much better than his prospective customers the real quality of his products. In Akerlof's famous Lemon Problem, the customers have no means to assertain the quality of products, which leads to a no-trade paradox [100]. When the customers are better equipped, optimal quality emerges [101–103]. One of the main issues is to understand under which conditions a manufacturer should diversify his production. Economics literature has approached this problem mainly with the help of utility functions. Several aspects have been studied, among them optimal quality-based product differentiation [104], firm competition by quality [105] and by price [106], the relation between product quality and market size [107], etc. (see [108, 109] for a review). We take the point of view of a monopolistic vendor faced to consumers deciding to buy one of his products according to their perception of its quality; using a probabilistic consumer choice framework makes it possible to avoid utility functions.

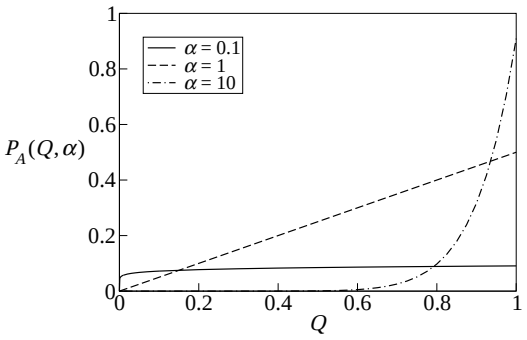


Figure 3.1: The acceptance probability $P_A(Q, \alpha)$ for various values of α.

3.1.1 Single product

We assume that the only difference between products lies in their quality $Q \geq 0$ which is therefore the main quantity of interest here. With a suitable choice of units, one can write the profit of a vendor per item sold as $1-F(Q)$ where F is an increasing function and $F(1) = 1$. For the sake of simplicity, we take $F(Q) = Q$; $F(Q) \propto Q^2$ is also found in the literature but does not alter qualitatively our results. While Q could in principle be greater than one, a vendor would never choose it, hence our analysis is restricted to $Q \in [0, 1]$ (this constraint will be relaxed by introduction of prices in Sec. 3.1.3).

We assume that a customer buys a product of quality Q with probability $P_A(Q)$. While there are many possible choices, *e.g.*those of Refs. [108, 110] or piece-wise linear functions as in Ref. [103], we shall mainly use

$$P_A(Q, \alpha) = \left(1 - \tfrac{1}{\alpha+1}\right) Q^\alpha \qquad (\alpha \geq 0) \tag{3.1}$$

where α is the acceptance parameter: for small α, P_A is mostly flat, resulting in a lack of quality discrimination; as α grows, the core of P_A shifts towards higher quality, which reflects enhanced discrimination abilities (see Fig. 3.1). We will use the shorthands "ignorant" for buyers with a small α and "informed" for those with a large α; an ignorant buyer is quite likely to reject even a perfect product.

If there are N buyers with acceptance parameters α_i ($i = 1, \ldots, N$), faced with a single product of quality Q, the vendor's expected profit X is

$$X(Q) = (1 - Q) \sum_{i=1}^{N} P_A(Q, \alpha_i) - Z. \tag{3.2}$$

where Z represents the fixed part of production costs due, for instance, to the initial investment needed to setup the manufacturing plant. Assuming that N is large, the fluctuations of X can be neglected. The structure of this expression

is similar to the profit function introduced in [108]. Since $X'(0) > 0$, $X'(1) < 0$, and $X(Q)$ is continuous, there is at least one Q maximizing X in $(0, 1)$.

In the following we adopt the point of view of the vendor and hence optimize his expected profit X.

Homogeneous population

When there is only one type of buyers, the expected profit simplifies to $X(Q) = N(1-Q)P_A(Q,\alpha) - Z$ which reaches its maximum at

$$Q^*(\alpha) = \frac{\alpha}{\alpha+1}. \tag{3.3}$$

Expectedly, $Q^*(\alpha)$ increases when the buyers have a sharper eye. The total optimal profit reads

$$X^*(\alpha) = N\frac{\alpha^{\alpha+1}}{(\alpha+1)^{\alpha+2}} - Z. \tag{3.4}$$

In Fig. 3.2 we report the expected optimal profit per customer $x^*(\alpha) := X^*(\alpha)/N$ as a function of α for $z := Z/N = 0.05$. When $z > 0$, a vendor only makes a profit when the quality is not too high or too low. Accordingly $x^*(\alpha)$ has a maximum at $\alpha_0 \approx 0.65$. Therefore, if the vendor cannot easily change Q, he should target a population with Q^*, or strive to modify the abilities of his prospective customers to detect quality, thereby increasing his profit. When $[0; \alpha_0)$, both the consumers and the vendor benefit from an increase in Q; we shall call it the cooperative region. Reversely, when $\alpha > \alpha_0$ the vendor suffers from excessive quality detection abilities of his customers; he could try a confusing marketing campain or rebranding so as to lower their abilities—this is the defensive region. A similar behaviour has been observed in [111]. In our case, the fact that the cooperative region is much smaller than the defensive region is a consequence of the shape of P_A. For instance, when the prefactor in $P_A(Q)$ changes from $1-(\alpha+1)^{-1}$ to $1-(\alpha+1)^{-1/3}$, the size of the cooperative region increases significantly.

Heterogeneous buyers

Heterogeneity brings in more surprises. Let us split the population into two groups, group $i = 1, 2$ consisting of N_i buyers with acceptance parameter α_i; the proportion of group i is denoted by $c_i := N_i/N$. The vendor's expected profit reads

$$X(Q) = N(1-Q)\big[c_1 P_A(Q,\alpha_1) + c_2 P_A(Q,\alpha_2)\big] - Z. \tag{3.5}$$

It is not possible to maximize X analytically. The result of numerical investigations is shown in Fig. 3.3 as a function of c_2 for $\alpha_1 = 0.1$ (ignorant buyers)

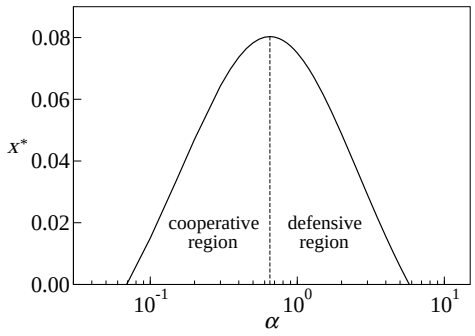


Figure 3.2: Optimal vendor's profit per buyer x^* as a function of α for $z = 0.05$.

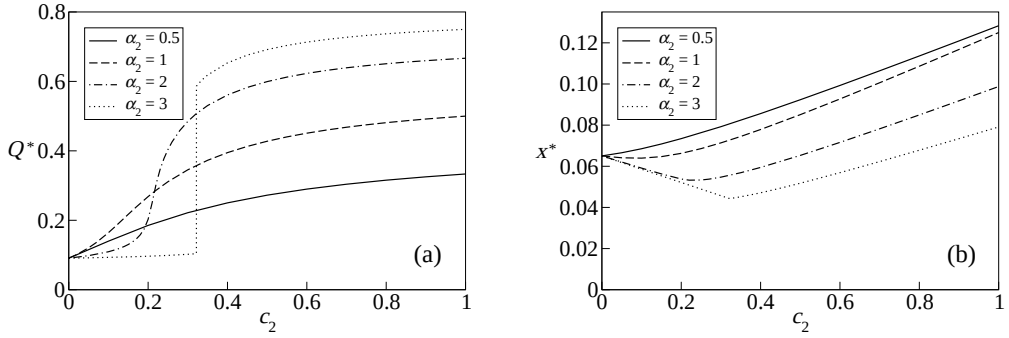


Figure 3.3: Optimal product quality Q^* (left) and vendor's optimal profit x^* (right) versus c_2 for various values of α_2; $\alpha_1 = 0.1$, $z = 0.01$.

and various choices of α_2. As expected, as the proportion of informed buyers increases, Q^* grows. But a surprising behaviour is found for instance when $\alpha_2 = 3$: at $c_2 \approx 0.32$ the optimal quality changes discontinuously. This is because $X(Q)$ has two local maxima. While for small c_2 the small-Q peak yields the largest profit, its relative height decreases as c_2 increases; accordingly the discontinuous transition occurs when the heights of the two maxima are equal. In Fig. 3.3 we also show the dependence of the optimal profit per buyer x^* on c_2. When group 2 has $\alpha_2 < \alpha_0$ (*e.g.*, $\alpha_2 = 0.5$), adding people with more demands regarding quality is beneficial to the vendor (Eq. (3.4)) and X is an increasing function of c_2. By contrast, when $\alpha_2 > \alpha_0$ the optimal profit first decreases as almost nobody of group 2 will buy anything and does so as long as group 2 has less influence on Q^* than group 1. Then group 2 supercedes group 1 and imposes its quality demands; the discussion generalises to an arbitrary number of groups. In other words, when society is too heterogeneous, it is impossible to satisfy all buyer groups with one product.

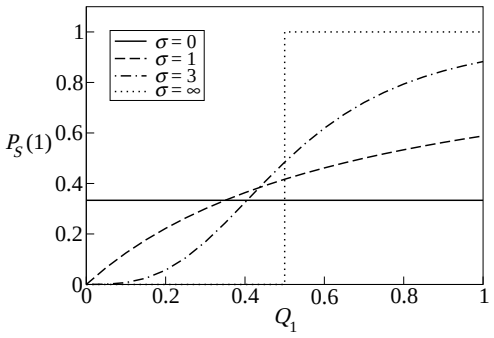


Figure 3.4: The probability to select variant 1 as a function of its quality Q_1 for various values of the selection parameter σ. In total three variants are displayed, the qualities $Q_2 = 0.5$ and $Q_3 = 0.2$ are fixed.

3.1.2 Multiple products

Now we assume that the vendor displays M product variants of different quality, at equal prices for the sake of simplicity, and that each buyer buys at most one item. A purchase is a two-step process, as a shopper has also to decide on a variant. The choice is also assumed to be probabilistic: variant $m = 1, \ldots, M$ is chosen according to

$$P_S(m|\boldsymbol{Q}, \sigma) = \frac{Q_m^\sigma}{\sum_{m'=1}^{M} Q_{m'}^\sigma}. \tag{3.6}$$

Here $\sigma \in [0; \infty)$ quantifies the selection ability of a given buyer. When σ is large, the buyer almost surely selects the best variant; on the contrary when $\sigma = 0$, $P_S(m) = 1/M$ for all m, *i.e.*, the buyer has no discerning power. Since P_S is normalized, each buyer purchases at most one item. Similar expressions appear in works on the influence of advertisement [110] and non-price competition [108], but other choices of functions would also be reasonable, such as exponentials as in the Logit model. All $Q_{m'}^\sigma$ have equal weight in Eq. (3.6); Sec. 3.1.2 generalizes this expression in order to take into account the proportions of displayed items. Finally, a more complete discussion on the plausibility of P_S is given in Sec. 3.1.4.

To summarize, the variant m with the quality Q_m is bought by buyer i with probability $P_S(m|\sigma_i, \boldsymbol{Q})P_A(Q_m, \alpha_i)$. As a consequence, if the vendor displays M variants to N buyers, his expected profit is

$$X(\boldsymbol{Q}) = \sum_{m=1}^{M}(1 - Q_m)\left(\sum_{i=1}^{N} P_S(m|\boldsymbol{Q}, \sigma_i)P_A(Q_m, \alpha_i)\right) - MZ. \tag{3.7}$$

This equation can be easily extended to account for special circumstances. For example, when Z is large, it may be profitable to produce one variant and achieve quality differentiation by artificially damaging a fraction of the production, *e.g.*by disabling some features [112]. In this case two variants with qualities

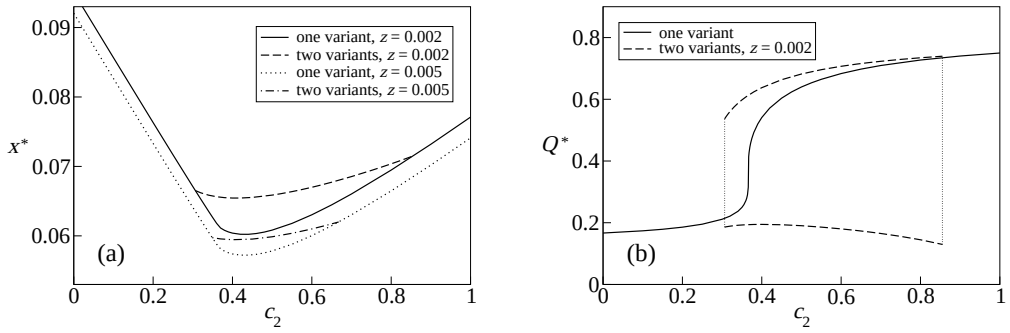


Figure 3.5: a) Optimal profits per buyer as a function of c_2; x_1^*: solid and dotted lines, x_2^*: broken and dashdot lines shown only when quality differentiation occurs. b) Optimal quality as a function of c_2, curves for Q_1^* and Q_2^* are only shown when $x_2^* > x_1^*$. Values of parameters are $\alpha_1 = 0.2$, $\alpha_2 = 3.0$, $\sigma_1 = 0.5$, and $\sigma_2 = 3.0$.

$Q_1 > Q_2$ are displayed but the profit per item sold is $1 - Q_1$ for both of them and the initial cost is reduced to Z.

Quality differentiation

For the sake of simplicity, we focus on two groups of customers consisting of N_i members with acceptance parameter α_i and selection power σ_i ($i = 1, 2$). The question is whether the vendor should display one or two products. In our framework, the answer is entirely determined by the respective optimal profit of each possibility, denoted by $X_1^*(Q)$ and $X_2^*(Q_1, Q_2)$.

Since manufacturing two products requires twice as much initial investment (by hypothesis), the region in which $X_2^* > X_1^*$ shrinks when z increases. This appears clearly in Fig. 3.5 where we plot the optimal profits versus $c_2 = N_2/N$ for two values of z. In addition, when $X_2^* > X_1^*$, the two optimal qualities Q_1^* and Q_2^* differ significantly. Quite clearly, the lower quality targets the group of ignorant buyers while the higher quality is for informed buyers. Remarkably, when $c_2 > 0.68$, the lower optimal quality is even smaller than the optimal quality $\alpha_1/(\alpha_1 + 1) = 1/6$ corresponding to a homogeneous population of ignorant customers. Notably, this downward distortion in a situation of a monopolistic vendor is also reported in [113].

Biased selection

Beyond $M = 2$, quality differentiation is equivalent to displaying more of the low-grade products. This suggests to introduce weights r_m ($\sum_{m=1}^{M} r_m = 1$) in the selection probability P_S of Eq. (3.6), taking into account for instance the effective visibility of each product due to advertisement or display position in

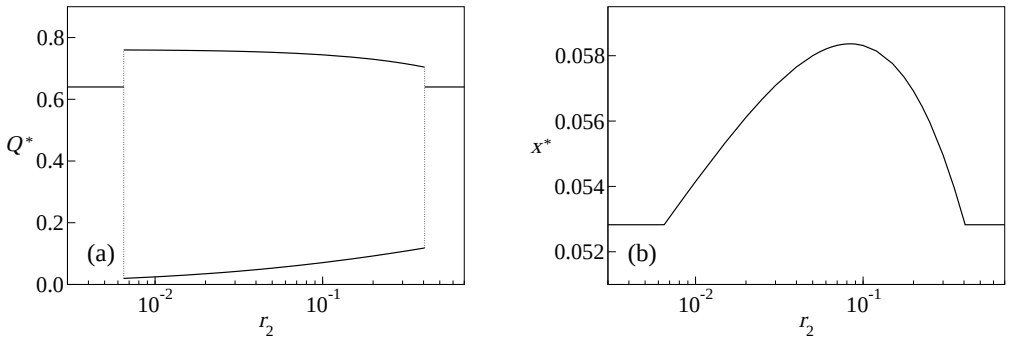


Figure 3.6: Optimal qualities (a) and the optimal profit (b) versus r_2 of the premium variant ($\alpha_1 = 0.2$, $\alpha_2 = 3$, $\sigma_1 = 0.2$, $\sigma_2 = 2$, $z = 0.01$, $c_2 = 0.5$, $M = 2$).

shops. The selection probability generalizes to

$$P'_S(m|\boldsymbol{Q}, \boldsymbol{r}, \sigma) = \frac{r_m Q_m^\sigma}{\sum_{m'=1}^{M} r_{m'} Q_{m'}^\sigma}. \tag{3.8}$$

Customers with high σ are able to pick the better product even when its effective proportion is small.

To study the effects of the proposed generalization, we use once again two groups of customers and choose the parameters so as to set the system in the quality differentiation region. Results of numerical optimization of the optimal profit are reported in Fig. 3.6, r_2 denotes the proportion of the premium variant. Differentiation occurs in a limited range of r_2: when r_2 is either too small or too large, buyers effectively notice only one variant and it is preferable for the vendor to produce only that one. In addition, x^* has a maximum at $r_2 \approx 0.08$, which comes from hiding the high quality variant to ignorant buyers while keeping it accessible to informed buyers.

3.1.3 Price

Let us now consider the price as a free parameter and investigate how the vendor should fix it optimally. Denoting the price by p, the profit per item is $p - Q$ which means that the maximum quality is $Q_{\max} = p$. In particular, if the vendor wishes to produce a better product that $Q_{\max}$, the price needs to be increased. The acceptance probability generalizes to $(\alpha \geq 0,\ p \in [Q; \alpha + 1])$

$$P_A(Q, p, \alpha) = \left(1 - \frac{p}{\alpha + 1}\right) \left(\frac{Q}{p}\right)^\alpha. \tag{3.9}$$

It satisfies two constraints: first, the higher the price, the smaller the acceptance probability. Second, because of the $p/(\alpha+1)$ term, the sensitivity towards prices

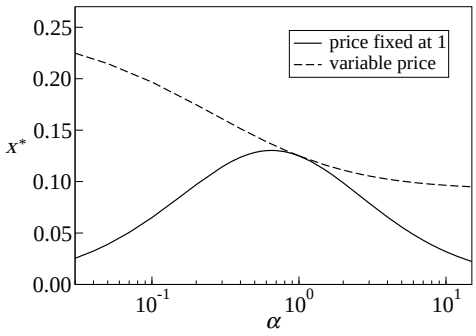


Figure 3.7: Optimal profit x^* vs acceptance parameter α: fixed price (solid line, the same curve as in Fig. 3.2) and variable price (dashed line), $z = 0$.

decreases as sensitivity to quality increases; similarly, quality must be judged with respect to price, hence the $(Q/p)^\alpha$ term. The discussion of the previous sections corresponds to $p = 1$.

We restrict our analysis to the simplest case of N identical buyers and one product. The expected profit reads

$$X_1(Q,p) = N(p-Q)\left(1-\frac{p}{\alpha+1}\right)\left(\frac{Q}{p}\right)^\alpha$$

with $p \leq \alpha + 1$ and $Q \in [0;p]$, it is maximized by

$$Q^* = \frac{\alpha}{2}, \quad p^* = \frac{\alpha+1}{2}. \tag{3.10}$$

Expectedly, the more informed the buyers, the better the products should be, but the vendor can charge a higher price. Because $p^* - Q^* = 1/2$ is a constant, there is no incentive in this model for exceptionally high prices for high quality variants. In Fig. 3.7, the resulting optimal profit per buyer x^* is shown together with the optimal profit when the vendor has fixed the price at 1. The liberty to set the price can increase the profit of the vendor quite considerably. The difference of profit for $\alpha > 1$ (informed buyers) is due to the fact that the vendor is allowed to charge a higher price for the high quality demanded by the buyers. By contrast, for $\alpha < 1$ the main improvement comes from the fact that $P_A(Q,p,\alpha)$ does not vanish when $Q \to 0$ and $p < 1$.

3.1.4 Discussion of the model plausibility

In order to better understand the need for both selection and acceptance procedures, it is worthwhile to consider some alternatives. One possibility to simplify our assumptions is to keep only the acceptance process with each displayed variant accepted or not according to the acceptance probability P_A. When M

variants with the qualities $Q_1, \ldots, Q_M$ are displayed, the probability P' that a given customer accepts at least one of them is

$$P'(Q_1, \ldots, Q_M) = 1 - \prod_{a=1}^{M} \big[1 - P_A(Q_a)\big]. \tag{3.11}$$

As M increases, P' converges to one. This means that the vendor can attract the buyers by displaying a large number of very bad products which is generally not the case. However, flooding of customers by low quality occurs under some special circumstances. This *economics of spamming* is briefly discussed in the next Appendix.

Another approach is to reduce the model to the best product selection governed by the selection probability P_S. Since this probability is normalized to one, each buyer surely buys one of the displayed variants and the vendor's profit maximization consequently yields zero quality. Obviously, such an optimal solution is pathological. One could eventually consider replacing the unity in the equation $\sum_{m=1}^{M} P_S(m|\boldsymbol{Q}, \sigma) = 1$ by an increasing function of the displayed qualities but this is effectively equivalent to our two step decision process. Thus we see that nor the selection step is sufficient to model the purchase process.

Finally, the generalization to diverse proportions of displayed variants, introduced in section 3.1.2, gives an additional argument. We see that while in the selection step both quality and proportion play their roles, in the final acceptance step it is only quality what matters. Thus these two steps are intrinsically different and attempts to merge them are artificial.

3.1.5 Conclusion

Due to the complexity of markets and human behaviour, attempts to propose a theory of the whole are illusory. However, simple models can bring insight to simple mechanisms at work in the real economy. Assuming probabilistic buyer behaviour, we formalized buyers' abilities, spanning from the zero information to the perfect information limits. Adopting the vendor's point of view, we examined the compromise between low quality which minimizes production costs and high quality which maximizes sales. In particular, the fact that customers are heterogeneous forces vendors to diversify their production. In other words, the large variety of products in free-market economies reflects in part the the information gathering and processing abilities of customers.

3.2 Selection by individual preferences

Truth is much too complicated to allow anything but approximations.

John von Neumann

The standard economics textbooks make the supply-demand law as one of the pillars of the modern economic theory. However, many people, especially economists (see for example [114]), gradually realize that the most important factor is missing in the traditional supply-demand law. The study of complex systems [115–117] has already led to novel approaches to market phenomena. In a previous work [111], a simple framework was introduced to treat both quality and information capability, yielding a generalized supply-demand law. It was assumed that a product is characterized by a single scalar variable: quality. However, in the modern economy we face a much more complex world where products have many attributes and consumers have heterogeneous tastes [103]. These preferences cannot be simply represented as price and quality alone. Therefore we generalize [111] to allow multiple variants of each product as well as many different tastes among consumers.

In consequence the producers face a dilemma: whether to target the average taste by producing a single or a few variants to leverage the economy of scale, or to match precisely each consumer's taste [118]. We shall see that the answer depends on the information level that the producers may access: whether they know, and how well they know the consumers' preferences. In addition, producers face also the nonlinear production costs. All the factors have to compromise to yield a combined result that gives various degrees of product diversity. With our approach, the supply-demand problem of producers with the capability of producing variations and consumers' diverse tastes becomes a matching problem [119, 120], where many mathematical and statistical mechanic tools are available to handle the complexity of the combinatorial problem.

To summarize, in this section we propose a simple market model and investigate its behavior under various circumstances. First we do not consider correlations between preferences of the parties included in the system. While unrealistic, this assumption allows us to discover basic properties of the model and outline the way of reasoning which can be used also in later, more realistic considerations. Then we discuss correlations and the ways how they can be introduced to the system. Finally we discuss how the system behavior changes under presence of correlations.

3.2.1 General framework: one vendor with many buyers

Let's begin with a market where only one vendor and M buyers are present. The vendor can produce N different variants of a product (*e.g.*, many different shoes). With regard to the market, he has to decide which variants it is optimal to produce. We assume that all buyers satisfied with the offer buy one item, others stay out of the trading. Buyers in the market we label with lowercase

Latin letters ($i = 1, \ldots, M$). The different variants the vendor can produce we label with Greek letters ($\alpha = 1, \ldots, N$). The price of variant α we label as P_α. We assume that every variant can be produced in as many pieces as it is needed and as fast as it is needed.

The simple structure sketched above offers us enough space to model basic features of real markets. To establish a mathematical model for the market we have to introduce some assumptions about participants' preferences and their consequences on the trading process. To keep complexity of the model at minimum we assume that buyer's opinion about a variant can by represented by one scalar quantity, which we call *cost* and label it with x; we assume $x \in [0; 1]$. The smaller is the cost $x_{i,\alpha}$, the bigger is the probability that buyer i is satisfied with variant α when asked. Preferences of the vendor are easier to introduce; they are represented by costs which he suffers during production and sale of a particular variant. The cost for variant α we label y_α and after a proper rescaling of monetary units $y_\alpha \in [0; 1]$. To simplify our considerations, we arrange the variants in order of cost: $y_1 < y_2 < \cdots < y_N$.

We stress a conceptual difference between vendor's and buyer's costs. The seller's cost y_α is strictly monetary—it represents a real amount of money (although in arbitrary units). In contrast, the buyer's cost $x_{i,\alpha}$ has no tangible interpretation, it simply represents something as airy as happiness with the given variant.

The vendor is able to produce N different variants. However, when he is producing more variants, his expenses grows due to need of an additional investment. The vendor's tendency to produce only few different variants can be modeled e. g. by a nonlinearity of expenses (doubled production of one single variant does not require doubled expenses). We adopt another approach; we assume that to initiate the production of a variant, the vendor has to pay additional charge $Z > 0$ (this represents initial costs).

Now let's assume that the vendor offered k most favorable variants (thus $\alpha = 1, \ldots, k$, $k \leq N$) to customers and the number of units sold of variant α is n_α. The total vendor's profit is

$$X(k, \{n_\alpha\}) = \sum_{\alpha=1}^{k} n_\alpha (P_\alpha - y_\alpha) - kZ. \tag{3.12}$$

Here the last term kZ comes for the initial costs of k produced variants, $P_\alpha - y_\alpha$ is the profit for one sold unit of variant α. Due to the monetary rescaling used to confine y_α to the range $[0; 1]$, units for profit, initial costs and prices are arbitrary.

It is natural to assume that when buyer i is asked about interest to buy variant α, the decision is based on the cost $x_{i,\alpha}$. We formalize this by the assumption that the probability of acceptance is a function of the variant cost; this function we call acceptance function. Obviously, $f(x)$ is a decreasing function of the cost

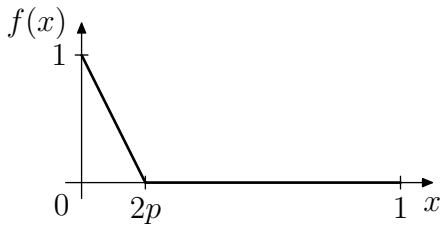


Figure 3.8: A particularly simple choice for the buyers' acceptance function $f(x)$.

x. Moreover, we assume $f(0) = 1$. This means that if a buyer considers a variant to be the perfect one, she surely buys it.

When we offer a random variant to one buyer, the acceptance probability is

$$\int_0^1 \pi(x) f(x) \mathrm{d}x \equiv p. \tag{3.13}$$

Here $\pi(x)$ is the probability distribution of cost x (*i.e.*, it defines what "to offer a random variant" really means). The probability p of accepting a random proposal is an important parameter of the model. From our everyday life we know that largely we do not agree to such an offer. For this reason we assume $p \ll 1$ in our calculations.

One example of a reasonable choice for the acceptance function is

$$f(x;p) = \begin{cases} 1 - x/2p & (0 \le x \le 2p), \\ 0 & (2p < x). \end{cases} \tag{3.14}$$

with $p < 0.5$ (see Fig. 3.8). This choice is especially convenient due to its simplicity. If we now assume the uniform distribution of the buyer's costs, $\pi(x) = 1$ for $0 \le x \le 1$, parameter p of the acceptance function Eq. (3.14) is just the probability p of accepting a random offer introduced in the previous paragraph.

In the rest of this section we assume that the prices of all variants are the same and equal to 1, $P_\alpha = 1$. This relieves us from many technicalities, and helps to highlight important features of the model. Nevertheless, generalization to various prices is straightforward.

3.2.2 No correlations in costs

We begin our investigation with the simplest case of the presented model—the market without correlations, where all costs y_α and $x_{i,\alpha}$ are mutually independent. We model this by costs uniformly distributed in the range $[0; 1]$. To keep variants ordered, we first draw their costs and then we renumber all variants achieve $y_1 < y_2 < \cdots < y_N$. It follows that after averaging over realizations, the formula $\langle y_\alpha \rangle = \alpha/(N+1)$ holds.

Vendor without knowledge of buyers' preferences

If a vendor wants to discover which variants are most acceptable for buyers, in a market without correlations each buyer has to be asked for preferences. This cannot be done in big markets, thus it is natural to investigate the case with no information about buyers' preferences on the vendor's side. In Sec. 3.2.2 we show that even an expensive global opinion survey brings only a negligible contribution to the vendor's income.

Without any information about preferences, the vendor is not able to discover which variants are most favored by buyers. Therefore the best strategy is to offer variants that are most favorable from his point of view. Let's label the number of variants the vendor is willing to offer as k. We assume that all these variants are available to buyers simultaneously, similarly to different types of shoes available in a shoe shop. Every buyer goes through the offered variants and decides whether some of them are suitable or not.

From the buyer's point of view, the vendor makes random proposals; the probability of accepting one particular offer is thus by definition equal to p. The probability P_A that one particular buyer accepts one of k proposed variants is complementary to the probability $(1-p)^k$ of denying all offered variants. Thus we have

$$P_A = 1 - (1-p)^k \approx 1 - \mathrm{e}^{-pk}, \tag{3.15}$$

where the approximation used is valid for $pk \ll 1$, *i.e.*for very choosy consumers (than p is a small quantity) and a small number of offered variants. Now the average number of items sold by the vendor to all M buyers is MP_A. Since no correlations are present, the average number of items sold of variant α is $\langle n_\alpha \rangle = MP_A/k$, it is a decreasing function of k.

The quantity of vendor's interest is the total profit X introduced in Eq. (3.12). Its expected value can be found using $\langle n_\alpha \rangle$, $\langle y_\alpha \rangle$, and P_A. We obtain

$$X_U(k) = M\big(1 - \mathrm{e}^{-pk}\big)\left(1 - \frac{1+k}{2(N+1)}\right) - kZ. \tag{3.16}$$

Here the subscript U reminds that we are dealing with an "uninformed" vendor. This function is sketched in Fig. 3.9 for three different choices of initial cost Z. The optimal number of variants the vendor should offer maximizes his profit. One can easily show that when $X'_U(0) < 0$, $X_U(k) < 0$ for all $k > 0$. Thus the condition $X'_U(0) < 0$, which can be rewritten as $Z > Mp$, characterizes a market where the optimal vendor's strategy is to stop the production and stay idle.

Since for every product numerous numerous variations can be made, the total number of variants the vendor can offer, N, is large. Thus we are allowed to assume that the optimal number of offered variants satisfies the condition $k_{\mathrm{opt}} \ll N$ and solve the maximization condition $X'_U(k) = 0$ approximately. We

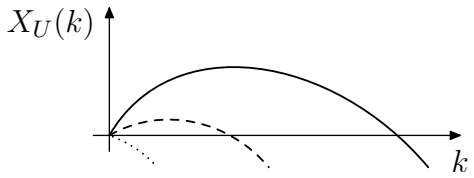


Figure 3.9: Expected profit of the vendor without informations, $X_U(k)$, drawn against k for small value of initial costs (solid line), medium initial costs (dashed line) and high initial costs (dotted line). In the last case the condition $Z > Mp$ is fulfilled and the optimal vendor's strategy is to stop the production.

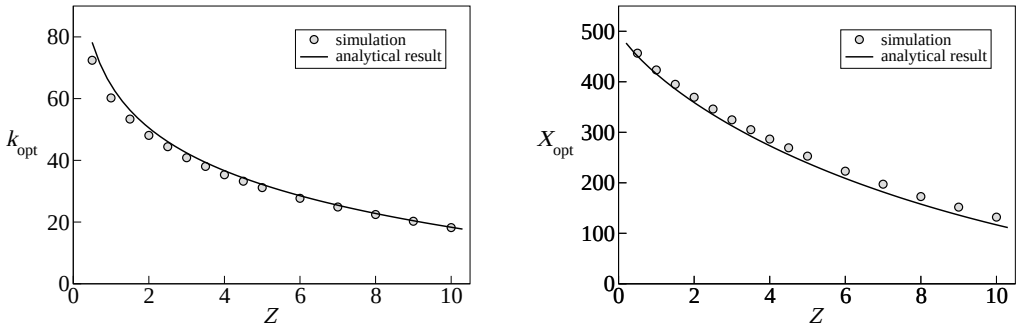


Figure 3.10: The optimal number of offered variants and the optimal profit as functions of initial cost Z for $M = 500$, $N = 2\,000$, and $p = 0.05$. Numerical results (empty circles) are averages over 1 000 realizations, analytical results (solid lines) come from Eq. (3.17). For the optimal profit arbitrary units are used.

obtain

$$k_{\rm opt} = \frac{1}{p}\,\ln\frac{Mp}{Z}, \quad X_{\rm opt} = M - \frac{Z}{p}\left(1 + \ln\frac{Mp}{Z}\right), \tag{3.17}$$

where $X_{\rm opt}$ is the optimal expected profit, $X_{\rm opt} = X_U(k_{\rm opt})$. The used approximations are valid when $Z \gg M/N$ and $p \ll 1$. In Fig. 3.10, these results are shown to match a numerical treatment of the problem. In the figure we see how the initial cost Z influences diversity of the vendor's production: decreasing Z increases differentiation of the vendor's supply in full agreement with expectations.

Improvement of the vendor's profit by a sequential offering of variants

So far we dealt with a very passive approach of the vendor. While offering k variants to the market, he had no influence on the sale. In consequence, due to the absence of correlations in the system, every offered variant had the same average number of items sold. In a big market this is a natural approach. While

the use of advertising can promote some variants, its treatment exceeds our scope.

In a small market a personal offering is possible. The vendor can promote favorable variants to increase the profit simply by offering the most favorable variant first. If a buyer is not interested, the second most favorable variant follows, etc. The average sale of the first variant is $\langle n'_1 \rangle = Mp$, for the second variant it is $\langle n'_2 \rangle = M(1-p)p$ and in general we have $\langle n'_\alpha \rangle = Mp(1-p)^{\alpha-1}$. Hence the expected total sale is

$$\sum_{i=1}^{k} Mp(1-p)^{i-1} = M\big[1-(1-p)^k\big].$$

This is equal to the expected total sale MP_A of the uninformed vendor in the previous section. We can conclude that the vendor's profit improvement (if any) does not come from an increased total sale but rather from an increased sale of the variants that are more profitable for the vendor.

Now we investigate the optimal number of variants to offer in this case, k'_{opt}. Since $\langle n'_\alpha \rangle$ decreases with α, at some moment it is not profitable to offer one more variant and the vendor's profit is maximized. The corresponding equation $\langle n'_\alpha \rangle = \langle n'_\alpha \rangle \langle y_\alpha \rangle + Z$ can be solved with respect to α, leading to k'_{opt}. When the total number of possible variants N is big, $\langle y_\alpha \rangle \ll 1$ and the term $\langle n'_\alpha \rangle \langle y_\alpha \rangle$ can be neglected. The approximate solution is then

$$k'_{\mathrm{opt}} \approx \frac{\ln(Z/Mp)}{\ln(1-p)}. \tag{3.18}$$

This optimal number of variants to offer is smaller than k_{opt} given by Eq. (3.17). We can also notice that when p is small, using approximation $\ln(1-p) \approx -p$ we are left with $k'_{\mathrm{opt}} \approx k_{\mathrm{opt}}$. This is an intriguing property—by the two different approaches we obtained the same result. To compare k'_{opt} in markets with different sizes, we plot it as a function of Z/M in Fig. 3.11. As can be seen, in a big market ($M \gtrsim 100\,000$) Eq. (3.18) fits well a numerical simulation of the system.

One can examine also the increase of the vendor's profit caused by the change of the sale method. Using previous results, the approximate formula $\Delta X_{\mathrm{opt}} \approx Z[1+\ln(Mp/Z)]/(Np^2)$ can be obtained. We see that when the total number of variants N is big, sequential offering results in a small growth of the vendor's profit. Nevertheless, in a system with a limited offer (small N) or with very choosy buyers (very small p), the improvement can be substantial.

Here we should notice, that the stopping condition "income greater than expenses" introduced above can be hard to use in practice. It is because n'_α is a random quantity and can drop to the disadvantageous region $n'_\alpha < Z + n'_\alpha y_\alpha$ even when $\langle n'_\alpha \rangle$ is big enough to cover the expenses. Thus for the vendor it is not enough to simply check profitability of the sale of one particular variant

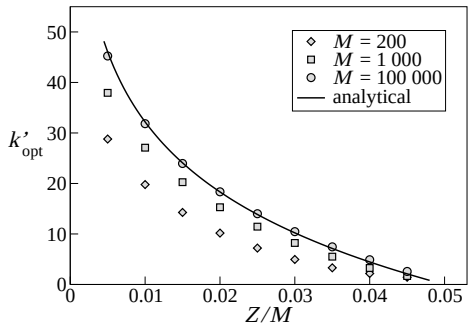


Figure 3.11: Successive offering: numerical and analytical results for the vendor using stopping condition described in the text in the markets with various sizes (on horizontal axis we have $q \equiv Z/M$). All numerical results are obtained as average of 10 000 realizations with $p = 0.05$, $N = 2\,000$; solid line represents Eq. (3.18).

n'_α. Rather he has to take into account sales of all previously offered variants. This is especially important in systems with a small number of buyers M where relative fluctuations are bigger. This effect is shown in Fig. 3.11 where numerical results for the vendor blindly using the stopping condition are shown for various market sizes. Clearly as M increases, numerical results approach the analytical result Eq. (3.18).

Competition of two vendors

In real markets we seldom find a monopolist vendor; competition and partition of the market is a natural phenomenon. To investigate the model behavior in such a case we introduce the second vendor to the market. We assume that the vendors differ by initial costs, which are Z_1 and Z_2. Again we do not consider the influence of advertisements and reputation, albeit they are vital in a market competition.

The course of the solution is similar to the one leading to Eq. (3.16). We label the number of variants offered by vendor 1 as k_1, the number of variants offered by vendor 2 as k_2, and we assume that there is no overlap between offered variants. The aggregate sale of two buyers is MP'_A where

$$P'_A = 1 - (1-p)^{k_1+k_2} \approx 1 - \exp[-p(k_1+k_2)].$$

With our assumptions about the equal status of the vendors, every offered variant has the same average sale. Therefore both vendors gain the share proportional to the number of variants they offer. Thus vendor 1 takes $k_1/(k_1+k_2$ of the total sale and vice versa. When $k_1/N, k_2/N \ll 1$, we can simplify the expected

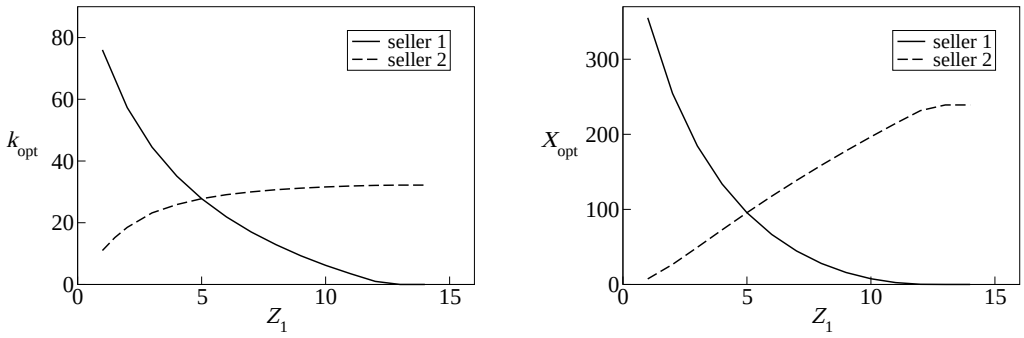


Figure 3.12: The optimal number of variants to offer (left) and the optimal profit (right) for vendor 1 (solid line) and for vendor 2 (dashed line) against Z_1. The initial cost of the second vendor is $Z_2 = 5.0$, $M = 500$, and $p = 0.05$.

profits to the form

$$X_1(k_1, k_2) = M\big(1 - \mathrm{e}^{-p(k_1+k_2)}\big)\,\frac{k_1}{k1 + k2} - k_1 Z_1, \tag{3.19a}$$

$$X_2(k_1, k_2) = M\big(1 - \mathrm{e}^{-p(k_1+k_2)}\big)\,\frac{k_2}{k1 + k2} - k_2 Z_2. \tag{3.19b}$$

Both parties maximize their profits by adjusting k_1 and k_2. The corresponding system $\partial_{k_1} X_1(k_1, k_2) = 0,\ \partial_{k_2} X_2(k_1, k_2) = 0$ cannot be solved analytically but its numerical treatment is straightforward. The result is shown in Fig. 3.12 where we have fixed the initial cost Z_2 to investigate how k_{opt} and X_{opt} for both vendors vary with Z_1.

We see that at $Z_1 \approx 12$ vendor 1 stops the production for he cannot stand the competition of vendor 2. By putting $k_2 = 0$ in the equations $\partial_{k_1} X_1(k_1, k_2) = 0$ and $\partial_{k_2} X_2(k_1, k_2) = 0$ we obtain the expression for the value Z_1^* when this price-out occurs

$$Z_1^* = \frac{Mp}{\ln(Mp/Z_2)}\left(1 - \frac{Z_2}{Mp}\right). \tag{3.20}$$

It is in a good agreement with the values found by a numerical simulation of the model. Important feature of this result is that it depends on the initial price Z_2 of the competitive vendor—decreasing the production costs can expel others from the market.

One can notice that when vendor 1 tries to increase the profit by deliberately increasing k_1 (with the intention to increase the sale), the term $-k_1 Z_1$ prevents the success of this strategy. As a result, the vendors have to adapt to each other. In mathematic terms, $X_1(k_{1\mathrm{opt}}, k_{2\mathrm{opt}}) \geq X_1(k_1, k_{2\mathrm{opt}})$. At the same time, the sum of profits is not maximized at $k_{1\mathrm{opt}}$ and $k_{2\mathrm{opt}}$. It is more profitable to remove the less efficient producer (the one with the higher value of initial costs). This is

an analogy of a real market where ruining (or taking over) of a competitor can improve company profit.

Informed vendor

Now we would like to investigate the artificial case of the market where the vendor knows costs $x_{i,\alpha}$ of all buyers. This knowledge can be used to increase the optimal profit. We start with a simpler question: if the vendor offers only one variant, how much the sale can be increased by a good choice of the variant?

The probability that buyer i is agreeable to buy variant α is $f(x_{i,\alpha})$. Since costs $x_{i,\alpha}$ are random and independent, only the average acceptance probability p plays a role and the number of users willing to buy this variant, n_α, is thus binomially distributed with the mean $\langle n_\alpha \rangle = Mp$ and the variance $\sigma^2 = Mp(1-p)$. When the number of buyers M is big, we can pass to a continuous approximation and assume the normal distribution of n_α

$$f(n_\alpha) \approx \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(n_\alpha - Mp)^2}{2\sigma^2}\right]. \tag{3.21}$$

The biggest value from the set $\{n_\alpha\}$ ($\alpha = 1, \ldots, N$) we label as m. This is the number of potential buyers for the most accepted variant and the vendor does the best when by offering this variant. The probability density $f_N(m)$ (often called *extremal distribution*) is

$$f_N(m) = \frac{N}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(m-Mp)^2}{2\sigma^2}\right] \left(\frac{1}{2} + \frac{1}{2}\,\mathrm{Erf}\left[\frac{m-Mp}{\sigma\sqrt{2}}\right]\right)^{N-1}. \tag{3.22}$$

The multiplication by N appears because we do not care which one of all N variants is "the most accepted" one and the error function term represents the probability that the remaining $N-1$ variants are less accepted.

Since we are interested in big values of N, we expect that the difference $\langle m \rangle - M$ is big in comparison with σ. Therefore we use the approximation $\mathrm{Erf}(x) \approx 1 - \exp[-x^2]/\sqrt{\pi x^2}$, which is valid for $x \gg 1$. When the error function value is close to one, we can also use the approximation $(1-x)^N \approx \exp[-Nx]$ ($x \ll 1$) to obtain

$$f_N(m) \approx \frac{N}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(m-Mp)^2}{2\sigma^2} - \frac{\sigma N}{\sqrt{2\pi}} \frac{\exp\left[-(m-Mp)^2/2\sigma^2\right]}{m-Mp}\right].$$

This form is too complicated to obtain an analytical result for $\langle m \rangle$. Instead we compute the most probable value $\tilde{m}$

$$\tilde{m} \approx Mp + \sigma\sqrt{2\ln\frac{\sigma^3 N}{\sqrt{2\pi}}}.$$

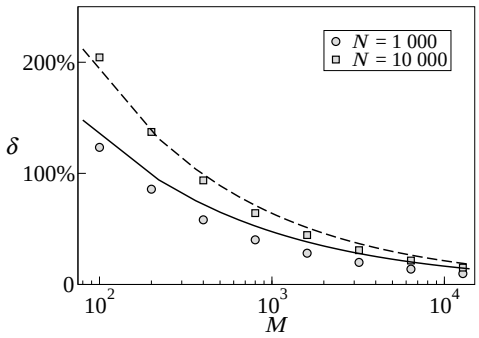


Figure 3.13: The relative growth of the vendor's sale δ is drawn against the total number of buyers M. Solid lines represent the analytical result, outcomes from numerical simulations are shown as symbols.

Here the first term Mp represents the average value of the sale and the additional term represents the gain arising from the additional vendor's knowledge. To get a better notion about the sale growth we use the relative sale growth

$$\delta \equiv \frac{\tilde{m} - Mp}{Mp} \approx \sqrt{\frac{1}{Mp} \ln \frac{pMN^2}{2\pi}}. \tag{3.23}$$

To simplify the formula, the assumption $p \ll 1$ has been used. A comparison of this result with a numerical simulation of the model is shown in Fig. 3.13. As can be seen, a good agreement is obtained.

When $\delta \ll 1$, all vendor's information is indeed useless and the average sale improvement is negligible. The inequality $\delta \ll 1$ leads to the condition

$$N^2 \ll \frac{2\pi}{Mp} \, \mathrm{e}^{Mp}. \tag{3.24}$$

Thus when the number of variants is not large enough, buyers' opinions in the uncorrelated market cannot be used to increase the vendor's sale and profit.

From the previous results we can draw useful implications about the vendor with perfect information, offering more than only one variant. When the total number of variants N is big, the number of variants offered by the vendor is small in comparison with N. Therefore the average sale of all offered variants is increased at most by δ given by Eq. (3.23) and the same applies to the total sale. However, the vendor is interested mainly in his profit. When we take into account different costs y_α of variants, the resulting growth of the income due to the informations is even smaller than δ because the variant with the highest sale can have a high cost for the vendor. Thus condition Eq. (3.24) is has more general consequences. It specifies the circumstances when even the perfect information about buyers' preferences do not help the vendor to achieve a significant improvement of his profit.

3.2.3 Correlations in the system

Now we would like to add one important flavor to the model—correlations. They arise from conformity of people's tastes (buyer-buyer correlations) and from the fact that high quality preferred by buyers results in high costs on the vendor's side (buyer-vendor anticorrelations). To approach the behavior of a real market, we investigate how these correlations influence our results obtained so far. Before doing so, we briefly discuss correlations from a general point of view.

Measures of correlations

A correlation is the degree to which two or more quantities are associated. We shall discuss different ways how to measure correlations and how to introduce them to the system. In particular, we would like to measure the correlation between two lists (vectors) of costs: $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ (two buyers) or $\boldsymbol{x}_i$ and $\boldsymbol{y}$ (a buyer and the vendor). All lists of our interest have length N and contain real numbers between 0 and 1. A common choice for the correlation measure is Pearson's correlation coefficient r. For lists $\boldsymbol{x}$ and $\boldsymbol{y}$ it is defined as

$$r^2 = \frac{\left[\sum_{\alpha=1}^{N}(x_\alpha - \overline{x})(y_\alpha - \overline{y})\right]^2}{\sum_{\alpha=1}^{N}(x_\alpha - \overline{x})^2 \sum_{\alpha=1}^{N}(y_\alpha - \overline{y})^2}. \tag{3.25}$$

This measure is sensitive to non-linear transformations of values in lists $\boldsymbol{x}$ and $\boldsymbol{y}$. In addition, since it originates in the least-square fitting of the data by a straight line, it measures only a linear correlation. For this reasons, we use a different correlation measure, Kendall's tau. For lists $\boldsymbol{x}$ and $\boldsymbol{y}$ it is given by the formula

$$\tau = \frac{2}{N(N-1)} \sum_{\alpha<\beta} \sigma_{\alpha\beta}, \qquad \sigma_{\alpha\beta} = \operatorname{sgn}\left[(x_\alpha - x_\beta)(y_\alpha - y_\beta)\right] \tag{3.26}$$

and it ranges from $+1$ (exactly the same ordering of lists $\boldsymbol{x}$ and $\boldsymbol{y}$) to -1 (reverse ordering of lists); uncorrelated lists have $\tau = 0$. Notably, Kendall's tau is insensitive to all monotonic mappings of the data. This is the strongest property we can expect from a correlation measure—more general transformations, nonmonotonic mappings, can sweep out any structure present in the data.

Lists with a given correlation degree

We would like to construct a set of lists that have mutual values of Kendall's tau equal to τ_0. Such a set would represent lists of buyers' preferences in an equally dispersed society. Since buyers' tastes are to a certain extent similar, we expect

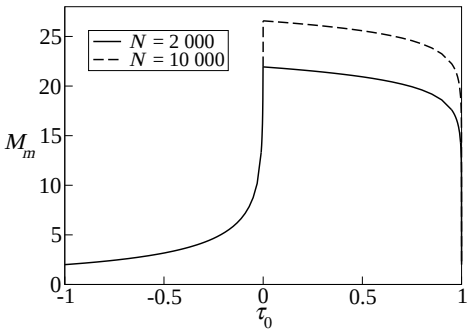


Figure 3.14: The upper bound M_m as a function of τ_0 for two different lengths of lists N. In both cases the upper bound M_m is the same over a large part of the region $\tau_0 < 0$ and drops to 1 when $\tau_0 \to 1$.

positive correlations with $\tau_0 > 0$. Nevertheless, in the following discussion we do not confine ourself to this region.

First we address a different question. Let's assume that between lists 1 and 2 there is τ_{12}, between lists 1 and 3 there is τ_{13}. Does it imply any constraints on τ_{23}? The answer is yes. It can be shown (see Sec. 3.2.5) that τ_{23} fulfills the inequality

$$|\tau_{12} + \tau_{13}| - 1 \leq \tau_{23} \leq 1 - |\tau_{12} - \tau_{13}|, \tag{3.27}$$

which is an analogy of the triangular inequality for side lengths of a triangle. From Eq. (3.27) we can draw various simple conclusions. First, if we want to construct three lists which have pairwisely τ_0, it is possible only for $-1/3 \leq \tau_0 \leq 1$.[2] Thus it is impossible to have more than two lists which are perfectly anticorrelated. Another simple result is that when $\tau_{12} = -1$, inevitably $\tau_{23} = -\tau_{13}$.

Now the question is whether we are able to create the whole system of M lists which all have pairwisely Kendall's tau equal to τ_0. It can be shown (see [121]) that the upper bound for M is

$$M_m = \begin{cases} 2 + \log_2 \frac{(1-\tau_0)(N-1)N}{4} & (\tau_0 \geq 0), \\ \min\left[2 + \log_2 \frac{(1-\tau_0)(N-1)N}{4},\ 2\log_2 \frac{1-\tau_0}{-\tau_0}\right] & (\tau_0 < 0). \end{cases} \tag{3.28}$$

As can be seen in Fig. 3.14, this quantity grows slowly with the list length N. Therefore to model a market with a large number of equally correlated buyers we would need an enormous number of possible variants.

[2] An example for lists with the pairwise value $\tau_0 = -1/3$: $x_1 = \{3, 2, 1\}$, $x_2 = \{2, 1, 3\}$ and $x_3 = \{1, 3, 2\}$.

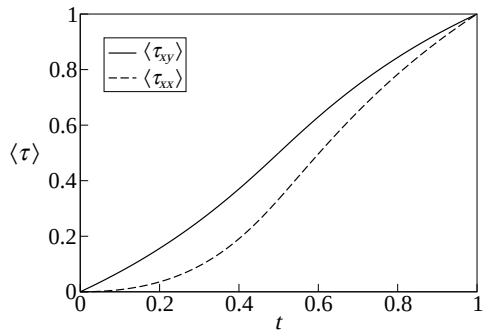


Figure 3.15: Kendall's tau vs. t for the correlated lists given by Eq. (3.29).

How to create correlated lists

In the previous paragraphs we found that the society with a fixed mutual correlation degree of buyers is limited in its size. Therefore to introduce correlations to the presented market model we need a different approach. While *copulas* represent a general tool (see *e.g.* [122, 123]), they are useful mainly for numerical simulations and offer only small possibilities for analytical results. Here we adopt a simpler way—the costs $\boldsymbol{x}_i$ and $\boldsymbol{y}$ are obtained by

$$x_{i,\alpha} = \frac{1}{2} + st\left(\frac{\alpha-1}{N-1} - \frac{1}{2}\right) + (1-t)\left(s_{i,\alpha} - \frac{1}{2}\right), \quad y_\alpha = \frac{\alpha-1}{N-1}, \tag{3.29}$$

where $s_{\alpha,j}$ is a random quantity distributed uniformly in the range $[0;1]$ (for other possibilities see [121]). The complicated form of the $x_{i,\alpha}$ has a simple meaning. The vendor's costs grow uniformly with α and buyers' costs are connected to the vendor's by the parameter $t \in [0;1]$. The term proportional to $1-t$ introduces a noise to the system, resulting in differences between buyers' and vendor's lists. Finally, the term $1/2$ represents the average value of buyers' costs. It is easy to check that $x_{i,\alpha}$ given by Eq. (3.29) is confined to the range $[0;1]$ for every $t \in [0;1]$ and $s = \pm 1$. The overall distribution of costs is uniform in the range $[0;1]$ and thus we avoid the problems of Eq. (3.29). Moreover, this construction is simple enough to tract the proposed model analytically.

Using the techniques shown in Sec. 3.2.5 we can find Kendall's tau in this case. In the limit $N \to \infty$ one obtains

$$\langle\tau_{xy}\rangle = \begin{cases} \frac{s}{6}\left(4u - u^2\right) & (u \le 1), \\ \frac{s}{6}\left(6 - \frac{4}{u} + \frac{1}{u^2}\right) & (u > 1), \end{cases}, \quad \langle\tau_{xx}\rangle = \begin{cases} \frac{u^2}{15}\left(10 - 6u + u^2\right) & (u \le 1), \\ \frac{1}{15}\left(15 - \frac{14}{u} + \frac{4}{u^2}\right) & (u > 1), \end{cases} \tag{3.30}$$

where again $u \equiv t/(1-t)$. Plots of $\langle\tau_{xy}\rangle$ and $\langle\tau_{xx}\rangle$ are shown in Fig. 3.15.

3.2.4 Uninformed vendor in the market with correlations

When we discussed the market without correlations, the probability distribution of the variant cost $\pi_\alpha(x_{i,\alpha})$ was independent of α. Consequently, the probability of accepting variant α

$$P_A(\alpha) = \int_D \pi_\alpha(x_{i,\alpha}) f(x_{i,\alpha}) \mathrm{d}x_{i,\alpha} \tag{3.31}$$

was also independent of α (we labeled $P_A \equiv p$). As a result, when we change the acceptance function $f(x)$ while preserving the quantity $\int_0^1 \pi(x) f(x) \mathrm{d}x$, the derived results remain unchanged. In the presence of correlations we witness a very different picture: the detailed shape of the acceptance function $f(x)$ is important.

To keep the algebra as simple as possible, from now on we adopt the simplest choice for $f(x)$: the step function $f(x) = 1 - \Theta(x - p)$. This means that a buyer accepts a proposed variant only when its cost is smaller than p. In addition we assume cost correlations created using Eq. (3.29). When the vendor has no information about the preferences of buyers, similarly to Sec. 3.2.2 the best strategy is to produce vendor's most favorable variants. First we focus on the case of positive correlations; in Eq. (3.29) we set $s = 1$ and $0 \leq t \leq 1$. Using Eq. (3.31) and the chosen step acceptance function $f(x)$, the probability that one buyer accepts variant α is

$$P_A(\alpha) = \begin{cases} 0 & 1 + (N-1)\,\frac{p}{t} < \alpha, \\ \frac{1}{1-t}\left[p - t\,\frac{\alpha - 1}{N-1}\right] & 1 + (N-1)\frac{p+t-1}{t} < \alpha < 1 + (N-1)\,\frac{p}{t}, \\ 1 & \alpha < 1 + (N-1)\frac{p+t-1}{t}. \end{cases} \tag{3.32}$$

Since we expect the total number of variants N to be very large and p rather small, the second region makes the major contribution and thus we simplify Eq. (3.32) to $P_A(\alpha) \approx (p - t\alpha/n)/(1-t)$.

When the vendor is simultaneously offering his k most favorable variants, the probability that a particular buyer denies all of them is

$$P_D(k) = \prod_{\alpha=1}^{k} \left[1 - P_A(\alpha)\right] = \prod_{\alpha=1}^{k} \left(1 - \frac{p}{1-t}\right)\left(1 + \frac{t\alpha}{N(1-t+p)}\right). \tag{3.33}$$

Since N is big, we use the approximation $1 - x \approx \exp[-x]$ to evaluate this expression analytically, leading to

$$\begin{aligned} P_D(k) &\approx \left(1 - \frac{p}{1-t}\right)^k \prod_{\alpha=1}^{k} \exp\left[-\frac{t\alpha}{N(1-t+p)}\right] \approx \\ &\approx \exp\left[-\frac{pk}{1-t} + \frac{tk^2}{2N(1-t+p)}\right]. \end{aligned} \tag{3.34}$$

Here we used also $1 - p/(1-t) \approx \exp[-p/(1-t)]$ which is valid when $p/(1-t)$ is small. When this is not the case, denying probability $P_D(k)$ approaches zero and thus accepting probability is virtually one regardless to the approximation used.

With respect to Eq. (3.29), the sum of expected vendor's costs can be written as

$$kZ + \sum_{\alpha=1}^{k} MP_S(\alpha)\frac{\alpha-1}{N-1} \approx kZ + \sum_{\alpha=1}^{k} MP_S(\alpha)\frac{\alpha}{N}. \tag{3.35}$$

Here the first term represents fixed costs for producing k different variants, $P_S(\alpha)$ is the probability that to one buyer variant α is sold. Since the probability that of the successful trade is $1 - P_D(k)$, from the condition $\sum_{\alpha=1}^{k} P_S(\alpha) = 1 - P_D(k)$ we can deduce

$$P_S(\alpha) = \frac{P_A(\alpha)}{\sum_{\alpha=1}^{k} P_A(\alpha)} \big[1 - P_D(k)\big]. \tag{3.36}$$

This corresponds to the portioning of the probability $1-P_D(k)$ among k variants according to their probability of acceptance.

Now we can use Eqs. Eq. (3.34), Eq. (3.35) and Eq. (3.36) to write down the expected profit of the vendor offering his k topmost variants $\overline{X}(k)$. It's not possible to carry out the maximization of this expression analytically—numerical techniques have to be used to find k_{opt} and X_{opt}. Results are shown in Fig. 3.16 as lines together with outcomes from a numerical simulation of the model; a good agreement is found for $st > 0$. Results confirm that positive correlations between buyers and the vendor increase the vendor's profit. This pattern is most obvious in the case $t = 1$ when the vendor can offer only the most favorable variant and still every buyer buys it.

In the numerical results shown in Fig. 3.16 we can notice one striking feature. When $st < 0$, k_{opt} changes rapidly and X_{opt} falls to zero quickly. Such a behavior is rather surprising for one do not expect abrupt changes in the region $st < 0$ when there were none in the opposite region $st > 0$. The reason for this behavior is simple—when $s = -1$, vendor's most preferred variants have cost too high to be accepted by a buyer. This effect can be quantified. When buyers' costs are generated by Eq. (3.29), the inequality $x_{i,\alpha} \geq t(N-\alpha)/(N-1)$ holds. Due to the acceptance function only the variants with cost smaller than p are accepted. Therefore only variants with $\alpha \geq \alpha_{\min}$ can be possibly accepted, where

$$\alpha_{\min} = 1 + (N-1)\,\frac{t-p}{t} \approx N(1-p/t). \tag{3.37}$$

Thus with negative correlations in the market, the vendor is able to sell the most favorable variant (the one with $\alpha = 1$) only if $p \geq t$. When $p < t$, the vendor sells no variants $\alpha = 1, \ldots, \alpha_{\min} - 1$. Since $\alpha_{\min}$ grows steeply with t (already

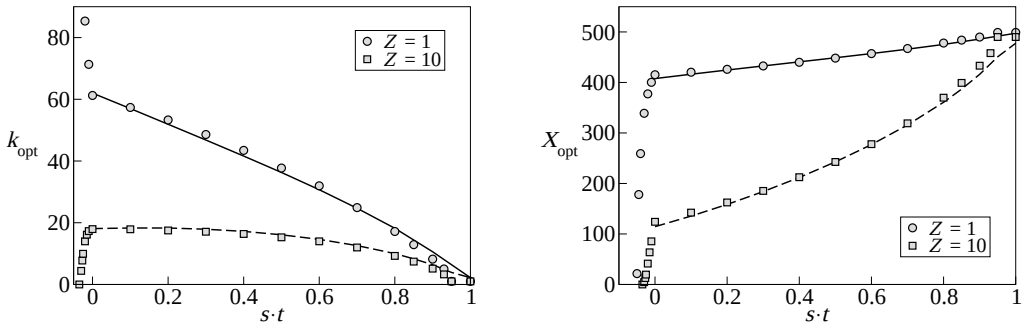


Figure 3.16: The optimal number of variants to produce (left) and the optimal profit drawn against st for two different values of the initial cost Z. Lines show analytical results derived above, symbols represent numerical simulations (averages of 1 000 realizations), model parameters are set to $N = 2\,000$, $M = 500$, $p = 0.05$. The decay of both quantities for $st < 0$ is in agreement with Eq. (3.37).

with $t = 2p$ one obtain $\alpha_{\min} = N/2$), the vendor offering his top k variants has to offer too many of them and he suffers both big initial costs and big costs y_α. As a result the vendor is pushed out of the market.

Without detailed investigation we can infer the system behavior when the step acceptance function is replaced by a different choice. In the limit of careless buyers with $f(x) = C$, the influence of correlations vanishes and both $k_{\rm opt}$ and $X_{\rm opt}$ do not depend on st and the model simplifies to the case investigated in Sec. 3.2.2. Thus as $f(x)$ gradually changes from the step function to a constant function, the dependence on st gets weaker. In particular, if the largest cost x for which $f(x) > 0$ is x_0 (for the step function $x_0 = p$), in Eq. (3.37) p is replaced by x_0. As a consequence, $\alpha_{\min} =$ decreases and the steep decline of $X_{\rm opt}$ in Fig. 3.16 shifts to a lower value of st.

3.2.5 Technicalities

Here we discuss two technical problems which are related to this section.

Proof of τ–inequality

Let's have three lists $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ consisting of N mutually different real numbers. The Kendall's τ for lists $\boldsymbol{x}$ and $\boldsymbol{y}$ can be written as $\tau_{xy} = (P_{xy} - N_{xy})/T$ where P_{xy} is the number of pairs $\alpha < \beta$ that satisfy $(x_\alpha - x_\beta)(y_\alpha - y_\beta) > 0$, N_{xy} is the same with a negative result of the product, and $T = N(N-1)/2$ is the total number of different pairs α, β. For the given values τ_{xy} and N it follows that

$$P_{xy} = T(1 + \tau_{xy})/2, \qquad N_{xy} = T(1 - \tau_{xy})/2. \tag{3.38}$$

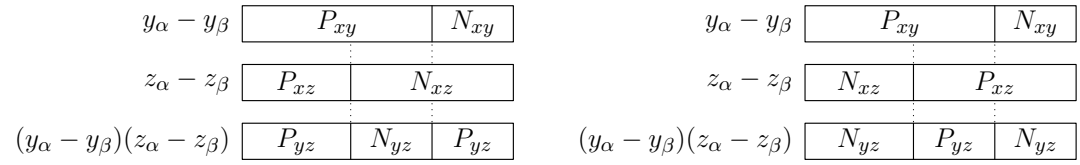


Figure 3.17: An illustration of the proof. The first case (left) has the biggest possible value of P_{yz}, the second case has (right) the smallest possible value of P_{yz}.

We would like to find bounds for τ_{yz} when τ_{xy} and τ_{xz} are given. First we reorder lists $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ so that lists $\boldsymbol{x}$ is sorted in the descending order and for $\alpha < \beta$ it is $x_\alpha - x_\beta > 0$. This rearrangement does not affect the values of $P_{xy}, P_{xz}, P_{yz}, N_{xy}, N_{xz}, N_{yz}$ and thus the values of Kendall's tau between lists remain also unchanged.

Since now all differences $x_\alpha - x_\beta$ are positive, from τ_{xy} we can deduce that there are P_{xy} positive differences $y_\alpha - y_\beta$ and N_{xy} negative differences. Similarly, P_{xz} differences $z_\alpha - z_\beta$ are positive and N_{xz} are negative. The values of P_{yz} and N_{yz} depend on the relative ordering of lists $\boldsymbol{y}$ and $\boldsymbol{z}$. The biggest possible value of P_{yz} occurs when positive differences $y_\alpha - y_\beta$ are aligned with positive differences $z_\alpha - z_\beta$ (see Fig. 3.17). By contrast, the smallest value of P_{yz} (and thus the smallest value of τ_{yz}) occurs when positive differences $y_\alpha - y_\beta$ are aligned with negative differences $z_\alpha - z_\beta$.

From Fig. 3.17 we see that P_{yz} and N_{yz} fulfill the inequalities

$$N_{yz} \geq |P_{xy} - N_{xz}|, \qquad P_{yz} \geq |P_{xy} - P_{xz}|.$$

Using $P_{ab} + N_{ab} = T$ and Eq. (3.38) we obtain

$$\begin{aligned} \tfrac{T}{2}|\tau_{xy} - \tau_{yz}| \leq & P_{yz} \leq T - \tfrac{T}{2}|\tau_{xy} + \tau_{yz}|, \\ -T + \tfrac{T}{2}|\tau_{xy} - \tau_{yz}| \leq & -M_{yz} \leq -\tfrac{T}{2}|\tau_{xy} + \tau_{yz}|. \end{aligned}$$

These two inequalities, summed together and divided by T, yield Eq. (3.27).

Expected values of $\langle\tau\rangle$

For lists created using Eq. (3.29) we can rearrange terms in Eq. (3.26) as follows

$$\langle\tau_{xx}\rangle = \frac{2}{N(N-1)} \sum_{\alpha<\beta} \langle\sigma_{\alpha\beta}\rangle = \langle\sigma_{\alpha\beta}\rangle.$$

Moreover, $\sigma_{\alpha\beta}$ can be rewritten as

$$\langle\sigma_{\alpha\beta}\rangle = P_{++} + P_{--} - P_{-+} - P_{+-} = 1 - 2P_{-+} - 2P_{+-} = 1 - 4P_{+-}.$$

Here P_{++} is the probability that both $x_\alpha - x_\beta$ and $x'_\alpha - x'_\beta$ are positive and so forth, the formulae $P_{+-} = P_{-+}$, $P_{++}+P_{--}+P_{-+}+P_{+-} = 1$ are used. According to Eq. (3.29) we write

$$x_\alpha - x_\beta = (1-t)(a_\alpha - a_\beta) + t(c_\alpha - c_\beta) \equiv (1-t)A + tC,$$
$$x'_\alpha - x'_\beta = (1-t)(b_\alpha - b_\beta) + t(c_\alpha - c_\beta) \equiv (1-t)B + tC$$

where A, B, C lie in the range $[-1; 1]$ and are equally distributed with the density $\varrho(A) = 1 - |A|$. Now we have ($t/(1-t) \equiv u$)

$$P(+-|C) = \begin{cases} \frac{1}{2}(uC)^2\left[1 - \frac{1}{2}(uC)^2\right] & (C \le 1/u), \\ 0 & (C > 1/u). \end{cases}$$

If $u \le 1$, the first case applies to all possible values of C, $P(+-|C) = 0$ is possible only if $u > 1$. Finally, using

$$P(+-) = \int_{-1}^{1} P(+-|C)\varrho(C)\,\mathrm{d}C$$

with $\varrho(C) = 1 - |C|$ it follows that

$$\langle \tau_{xx} \rangle = \begin{cases} \frac{u^2}{15}\left(10 - 6u + u^2\right) & (u \le 1), \\ \frac{1}{15}\left(15 - \frac{14}{u} + \frac{4}{u^2}\right) & (u > 1). \end{cases}$$

The quantity $\langle \tau_{xy} \rangle$ can be derived in the same way.

The variance of τ_{xx} can be found by a direct computation of $\langle \tau_{xx}^2 \rangle$. We have

$$\tau_{xx}^2 = \frac{1}{N^2(N-1)^2}\left(\sum_{\alpha\neq\beta} \sigma_{\alpha\beta}^2 + \sum_{\alpha\neq\beta}\sum_{\gamma\neq\delta} \sigma_{\alpha\beta}\sigma_{\gamma\delta}\right).$$

The averaging procedure is straightforward. At the end we obtain

$$\sigma_\tau^2 = \langle \tau_{xx}^2 \rangle - \langle \tau_{xx} \rangle^2 \approx \frac{4}{N}\left(\langle \sigma_{\alpha\gamma}\sigma_{\gamma\beta} \rangle - \langle \sigma_{\alpha\beta} \rangle^2\right),$$

where terms proportional to higher powers of $1/N$ were neglected. The variance is largest when $t = 0$, for $t = \pm 1$ obviously $\sigma_\tau = 0$.

3.2.6 Conclusion

In this section we proposed a simple market model. While accessible to analytical solutions, it exhibits many features of real markets—diversification of the vendor's production and market competition are used as examples. The diversification is presented as an interplay between the vendor's pursuit to follow the

buyers' tastes and the costs growing with the number of produced variants. We also show that in a market with many buyers without preferences correlations, the knowledge of these preferences doesn't increase the vendor's profit. When correlations are introduced to the system, many technical complications arise. Nevertheless, the results are consistent with the expectations: a positive correlation between the buyers' and vendor's costs improves the vendor's profit. Also, when interests of the two parties diverge (the correlation are negative), the vendor is able to make only a small or even no profit.

4 Entropy, information and portfolio optimization

I was seldom able to see an opportunity until it had ceased to be one.

Mark Twain

Portfolio optimization is one of the key topics in finance. It can be characterized as a search for a satisfactory compromise between maximization of the investor's capital and minimization of the related risk. The outcome depends on properties of the investment opportunities and on the investor's attitude to risk but crucial is the choice of the optimization goals. The problem has been pioneered by Markowitz in [124], where the Mean-Variance efficient portfolio has been introduced: it minimizes the portfolio variance for any fixed value of its expected return. Since this rule can be only justified under somewhat unrealistic assumptions (namely a quadratic utility function or a normal distribution of returns, in addition to risk aversion), it should be considered as a first approximation of the optimization process. Later, several optimization schemes inspired by Markowitz's work has been proposed [125–127]. For a recent thorough overview of the portfolio theory see [128, 129].

A different perspective has been put forward by Kelly in [130], where he shows that the optimal strategy for the long run can be found by maximising the expected value of the logarithm of the wealth after one time step. The optimality of this strategy has long been treated and proven in many different ways [131–133] and, according to [134], it can be successfully used in real financial markets. Although the Kelly criterion does not employ a utility function, as pointed out by the author himself, a number of economists have adopted the point of view of utility theory to evaluate it [135–138]. Recently, the superiority of typical outcomes to average values has been discussed from a different point of view in [139, 140]. A generalization of the Kelly's ideas led to the universal portfolios proposed in [141].

The original concept of Kelly focuses on a single investment in many successive time periods. In section 4.1 we briefly discuss this original problem and present main results. In section 4.2 we generalize the classic Kelly game by introducing simultaneous risky games and risky games with unknown proper-

ties. These extensions are well suited for investigating the effects of diversification and limited information on investment performance. However, in complex models of real investments, important features can get unnoticed. Therefore we replace realistic assumptions about the available investment opportunities (*e.g.*, a log-normal distribution of returns) by simple risky games with binary outcomes. In section 4.3 we derive various analytical results for a more realistic case when investment returns have a log-normal distribution. Finally, in section 4.4 we propose a new quantity to characterize the influence of correlations on investment performance and examine it both in artificial and real conditions.

4.1 Short summary of the Kelly game

> *In the long-run, thus defined, a penny invested at* 6.01% *is better—eventually becomes and stays greater—than a million dollars invested at* 6%.
>
> *Harry Markowitz*

In the original Kelly game, an investor (strictly speaking, a gambler) with the starting wealth W_0 is allowed to invest a part of the available wealth in a risky game and this investment is repeated many times. In each turn, the risky game has two possible outcomes: with the probability p the stake is doubled, with the complementary probability $1-p$ the stake is lost. The winning probability p is constant and known to the investor, the investor neither consumes nor adds new cash to his portfolio.

We introduce the game return R which is defined as $R := (W_r - W_i)/W_i$ where W_i is the invested wealth and W_r is the resulting wealth. For the risky game described above the possible returns per turn are $+1$ (win results in $W_r = 2W_i$) and -1 (loss results in $W_r = 0$). Since properties of the risky game do not change in time, the investor bets the same fraction f of the actual wealth in each turn. The wealth then follows a multiplicative process, after N turns it is equal to

$$W_N(R_1, \ldots, R_N) = W_0 \prod_{i=1}^{N} (1 + f R_i) \tag{4.1}$$

where R_i is the game return in turn i. If we label the number of winning turns as w, it simplifies to the form

$$W_N(w) = W_0 \, (1+f)^w (1-f)^{N-w}. \tag{4.2}$$

Since the successive returns R_i are independent, from Eq. (4.1) the average wealth after N turns can be written as (averages over realizations of the risky game we label as $\langle \cdot \rangle$)

$$\langle W_N \rangle = W_0 \, \langle 1 + f R_i \rangle^N = W_0 \left[1 + (2p-1) f \right]^N. \tag{4.3}$$

Maximization of $\langle W_N \rangle$ can be used to optimize the investment. Since for $p < 1/2$, $\langle W_N \rangle$ is a decreasing function of f, the optimal strategy is to refrain from investing: $f^* = 0$. By contrast, for $p > 1/2$ the quantity $\langle W_N \rangle$ increases with f and thus the optimal strategy is to stake everything in each turn: $f^* = 1$. Then, while $\langle W_N \rangle$ is maximized, the probability of a bankrupt in first N turns is $1 - (1-p)^N$. Thus in the limit $N \to \infty$, the investor bankrupts inevitably. Thus, maximization of $\langle W_N \rangle$ is not a good criterion for a long run investment.

In his seminal paper [130], Kelly suggested maximization of the exponential growth rate of the investor's wealth

$$G = \lim_{N\to\infty} \frac{1}{N} \log_2 \frac{W_N}{W_0} \tag{4.4}$$

as a criterion for investment optimization; without affecting results, we use natural logarithm instead. Due to the multiplicative character of W_N, G can be rearranged as

$$G = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \ln\left(1 + f R_i\right) = \langle \ln W_1 \rangle, \tag{4.5}$$

Notice that while we investigate investments within a long time horizon, the wealth W_1 after turn step plays a prominent role in the optimization. For the risky game introduced above is $\langle \ln W_1 \rangle = p \ln(1+f) + (1-p)\ln(1-f)$ which is maximized by the investment fraction

$$f_{\mathrm{K}}(p) = 2p - 1. \tag{4.6}$$

When $p < 1/2$, $f_{\mathrm{K}} < 0$ (a short position) is suggested. Here we allow only non-negative investment fractions and thus for $p < 1/2$ the optimal investment is $f_{\mathrm{K}} = 0$. For $p \geq 1/2$, the maximum of G achieved by f_{K} can be rewritten as

$$G_{\mathrm{K}}(p) = \ln 2 - S(p), \tag{4.7}$$

where $S(p) = -\big[p \ln p + (1-p)\ln(1-p)\big]$ is the entropy assigned to the risky game with the winning probability p. The additional term $\ln 2$ is due to our choice of the basis of the logarithms.

If we define the compounded return per turn R_N by the formula $W_N = W_0\,(1 + R_N)^N$ and its limit value as $\mathcal{R} := \lim_{N\to\infty} R_N$, it can be shown that $\mathcal{R} = \exp[G] - 1$. Thus maximization of $\mathcal{R}$ leads again to Eq. (4.6). While G is usually easier to compute than $\mathcal{R}$, in our discussions we often use $\mathcal{R}$ because it is more illustrative in the context of finance. Hence the maximum of $\mathcal{R}$ can be of interest; using $\mathcal{R}_{\mathrm{K}} = \exp[G_{\mathrm{K}}] - 1$ we get

$$\mathcal{R}_{\mathrm{K}}(p) = 2p^p (1-p)^{1-p} - 1. \tag{4.8}$$

When $p = 1/2$, $\mathcal{R}_{\mathrm{K}} = 0$; when $p \to 1$, $\mathcal{R}_{\mathrm{K}} = 1$.

There is one more way to f_{K}. If the winning probability is p, in a large number of turns N there will be on average Np winning and $N(1-p)$ loosing turns. Since in a winning turn wealth is multiplied by $1+f$ and in a loosing turn by $1-f$, one can write the typical wealth after N turns as

$$W_N^{\mathrm{typ}} = W_0\,(1+f)^{Np}(1-f)^{N(1-p)}. \tag{4.9}$$

For a given p, this expression is maximised by $f^* = 2p-1$. Thus, Kelly's criterion can be restated as maximization of the typical outcome.

The results obtained above we illustrate on a particular risky game with the winning probability $p = 0.6$. Since $p > 0.5$, it is a profitable game and a gambler investing all the available wealth has the expected return $\langle R\rangle = 2p - 1 = 20\%$ in one turn. However, according to Eq. (4.6) in the long run the optimal investment fraction is $f_{\mathrm{K}} = 0.2$. Thus, the expected return in one turn is reduced to $0.2 \times 20\% = 4\%$. For repeated investment, the average compounded return $\mathcal{R}$ is a better quantity to measure the investment performance. From Eq. (4.8) it follows that in this case it is equal to 2.0%. We see that a wise investor gets in the long run much less than the illusive return 20% of the given game (and a naive investor gets even less). In the following section we investigate how diversification (if possible) can improve this performance.

After discussion of basics we are ready to put Kelly's criterion to a somewhat wider perspective. It can be said that the Kelly-optimal portfolio offers a compromise between trend following (an asset with a long term good performance gets a large share of the optimal portfolio) and trend reversing (price increases lead to selling of shares). This description applies also to Cover's universal portfolios which were obtained as a generalization of the Kelly's approach.

4.2 Diversification and limited information in the Kelly game

All models are wrong, but some are useful.

George Box

In this section we analyse three modifications of the original Kelly game. First we assume that instead of one gam played in many rounds, there are several games which can be played simultaneously. The obtained results we use to compare the outcomes of outsider who maximise their gains by diversification and insiders who maximise their gains by finding additional information about the available investment opportunities. In the last part we waive the assumption that the game properties (*e.g.*, the winning probability p) are exactly known to the investor—we investigate what happens when they need to be inferred from past outcomes.

4.2.1 Simultaneous independent risky games

Now we generalize the original Kelly game assuming that there are M independent risky games which can be played simultaneously in each time step (correlated games will be investigated in a separate work). In game i ($i = 1, \ldots, M$) the gambler invests the fraction f_i of the actual wealth. Assuming fixed properties of the games, this investment fraction again does not change in time. For simplicity we assume that all games are identical, *i.e.*in each one with the probability p the invested wealth is doubled and with the probability $1 - p$ the investment is lost. Thus the return R_i of game i has two possible values, -1 and $+1$. Since the available games are identical, the optimal investment fractions are equal—we set $f_i = f$. Thus, the investment optimization is simplified to a one-variable problem.

Notice that with the given set of risky games, there is the probability $(1-p)^M$ that in one turn all M games are loosing. In consequence, for all $p < 1$ the optimal investment fraction f^* is smaller than $1/M$ and thus $Mf^* < 1$ (otherwise the gambler risks getting bankrupted and the chance that this happens approaches one in the long run). Thus, there is no incentive for borrowing additional money for the investment.

If in one turn there are w winning and $M - w$ loosing games, the investment return is $(2w - M)f$ and the investor's wealth is multiplied by the factor $1 + (2w - M)f$. Consequently, the exponential growth rate is

$$G = \langle \ln W_1 \rangle = \sum_{w=0}^{M} P(w; M, p) \ln \big[1 + (2w - M)f\big], \tag{4.10}$$

where $P(w; M, p) = \binom{M}{w} p^w (1-p)^{M-w}$ is a binomial distribution. The optimal investment fraction is obtained by solving $\partial G/\partial f = 0$. If we rewrite $2w - M = [f(2w - M) + 1 - 1]/f$ and use the normalization of $P(w; M, p)$, we simplify this equation to

$$\sum_{w=0}^{M} \frac{P(w; M, p)}{1 + (2w - M)f} = 1. \tag{4.11}$$

For $M = 1$ we obtain the well-known result $f_1^* = 2p - 1$, for $M = 2$ the result is $f_2^* = (2p - 1)/(4p^2 - 4p + 2)$. Formulae for $M = 3, 4$ are also available but too complicated to present here; for $M \geq 5$, Eq. (4.11) has no closed solution. This motivates us to investigate analytical approximations of the optimal investment fraction f^*. In complicated cases where such approximations perform badly, numerical algorithms are still applicable [142].

Approximate solution for an unsaturated portfolio

By an unsaturated portfolio we mean the case when a small part of the available wealth is invested, $Mf^* \ll 1$. Then also $|(2w - M)f^*| \ll 1$ and this allows us to

use the expansion $1/(1+x) \approx 1 - x + x^2 \pm \ldots$ ($|x| \ll 1$) in Eq. (4.11). If we take only the first three terms into account, we obtain $\sum_{w=0}^{M} P(w; M, p)\big[1 - f(2w - M) + f^2(2w - M)^2 \pm \ldots\big] = 1$. After the summation, the approximate formula for the optimal investment fraction follows

$$f^*(p) = \frac{2p-1}{M(2p-1)^2 + 4p(1-p)}. \tag{4.12}$$

When $p - 1/2 \ll 1/M$, this result simplifies to $f^* = 2p - 1$; this means that the gambler invests in each game as if other games were not present. Notice that for $M = 1, 2$, Eq. (4.12) is equal to the exact results obtained above.

Approximate solution for a saturated portfolio

By a saturated portfolio we mean the case when almost all available wealth is invested, $1 - Mf^* \ll 1$. The extreme is achieved for $p = 1$ when the solution is $f^* = 1/M$ as all wealth is distributed evenly among the games. We introduce the new variable $x := 1/M - f$ and rewrite Eq. (4.11) as

$$\frac{P(0; M, p)}{xM} + \sum_{w=1}^{M} \frac{P(w; M, p)}{2w/M - x(2w - M)} = 1.$$

Since according to our assumptions $0 < x \ll 1/M$, to obtain a leading order approximation for f^* we neglect x in the sum which is then equal to $\frac{M}{2}\langle 1/w\rangle$. The crude approximation $\langle 1/w\rangle \approx 1/\langle w\rangle$ leads to the result

$$f^* = \frac{1}{M}\left[1 - \frac{2p(1-p)^M}{2p-1}\right]. \tag{4.13}$$

Notice that in the limit $p \to 1$ we obtain $f^* = 1/M$ as expected.

Approximations Eq. (4.12) and Eq. (4.13) can be continuously joined if for $p \in [\frac{1}{2}; p_c]$ the first one and for $p \in (p_c; 1]$ the second one is used. The boundary value p_c is determined by the intersection of these two results. In Fig. 4.1, a comparison of the approximations with numerical solutions of Eq. (4.11) is shown and a good agreement can be seen for most parameter values. As Fig. 4.1a shows, the largest deviations appear for a mediocre number of games ($M \simeq 5$) and a mediocre winning probability ($p \simeq p_c$).

4.2.2 Diversification vs information

In real life, investors have only a limited information about the winning probabilities of the available risky games. These probabilities can be inferred using historical wins/losses data but these results are noisy and the analysis requires investor's time and attention (we investigate it in detail in Sec. 4.2.3). At the

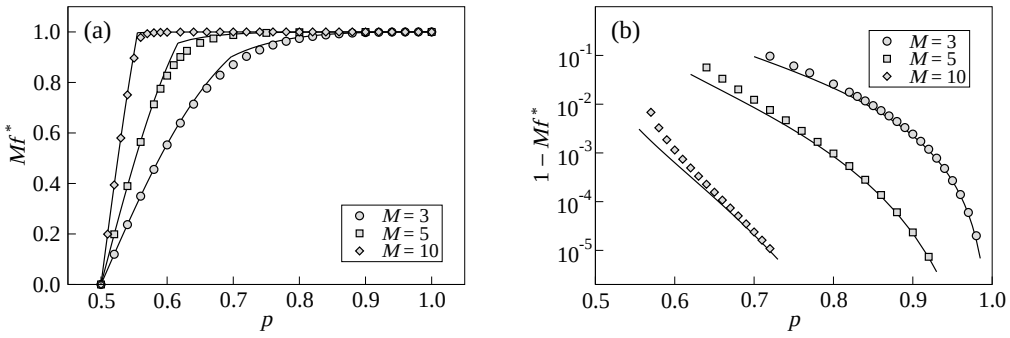


Figure 4.1: The comparison of numerical results for the optimal investment fraction f^* (shown as symbols) with the analytical results given in Eq. (4.12) and Eq. (4.13) (shown as solid lines). (a) The total investment Mf^* as a function of p. (b) To judge better the approximation for a saturated portfolio, the univested fraction $1 - Mf^*$ is shown as a function of p.

same time, insider information can improve the investment performance substantially. A similar insider-outsider approach can be seen in the classical paper on efficient markets [143] and in a simple trading model [144]. We try to model the described situation by a competition of two investors who can invest in multiple risky games; each of these risky games has the winning probability alternating randomly between $p + \Delta$ and $p - \Delta$ (with even odds, $p > 1/2$, $\Delta \leq 1 - p$). The insider focuses on one game in order to obtain better information about it—we assume that the exact winning probability is available to him. By contrast, the outsider knows only the time average p of the winning probability. The lack of information he tries to compensate by investing in multiple risky games. Both strategies have their advantages, below we investigate when the outsider performs better than the insider.

The insider knows the winning probability in each turn and thus the optimal investment fraction is given by Eq. (4.6). If $p - \Delta > 1/2$, the insider invests in each turn. If $p - \Delta \leq 1/2$, the insider invests only in those turns when the winning probability is $p + \Delta$ which means that the risky game is profitable. Combining the previous results, the exponential growth rate of the insider $G_I = \langle \ln W_1 \rangle$ can be simplified to

$$G_I = \begin{cases} \frac{1}{2}\big[\ln 2 + S(p+\Delta)\big] & p - \Delta \leq 1/2, \\ \frac{1}{2}\big[\ln 2 + S(p+\Delta)\big] + \frac{1}{2}\big[\ln 2 + S(p-\Delta)\big] & p - \Delta > 1/2, \end{cases} \tag{4.14}$$

where $S(p)$ is the entropy of the risky game with the winning probability p as in Eq. (4.7). About the outsider we assume that he invests in M identical and independent risky games. From his point of view, each risky game is described by the average winning probability p. Consequently, the exponential growth

rate of his investment is given by Eq. (4.10) and for the optimal investment fraction results from the previous section apply.

Now we can find the limiting values of the amplitude Δ above which the insider performs better that an outsider investing in two or more games. That is, we want to solve the equation

$$G_I(p,\Delta) = G_O(p,M) \tag{4.15}$$

with respect to Δ. Although the outsider's optimal investment fraction can be found analytically for $M \leq 4$, due to the form of $G_I(p,\Delta)$ it is impossible to obtain an analytical expression for Δ. An approximate solution can be derived by expanding $G_I(p,\Delta)$ in powers of Δ. First terms of this expansion has the form

$$G_I(p,\Delta) = \begin{cases} \frac{1}{2}\left[G_{\mathrm{K}}(p) + \Delta\big(\ln p - \ln[1-p]\big) + \frac{\Delta^2}{2}\left(\frac{1}{p} + \frac{1}{1-p}\right)\right] & p-\Delta \leq 1/2, \\ G_{\mathrm{K}}(p) + \frac{\Delta^2}{2}\left(\frac{1}{p} + \frac{1}{1-p}\right) + \frac{\Delta^4}{12}\left(\frac{1}{p^3} + \frac{1}{(1-p)^3}\right) & p-\Delta > 1/2. \end{cases}$$

By substituting this to Eq. (4.15) we obtain a quadratic (when $p-\Delta \leq 1/2$) or biquadratic (when $p-\Delta > 1/2$) equation for Δ which can be solved analytically. In this way we get $\Delta(p,M)$; when Δ is larger than $\Delta(p,M)$, at given p and Δ the insider performs better than an outsider diversifying into M games. When $p-1/2 \ll 1$, using expansions of $G_I(p,\Delta)$ and $G_O(p,M)$ the lowest order approximation of Δ is $\Delta(p,M) = (p-1/2)(\sqrt{2M}-1)$. To review the accuracy of our approximations, in Fig. 4.2 analytical results are shown for $M = 2,3,4$ together with numerical treatment of Eq. (4.15). Notice that the proposed approximation works better for $M = 2$. This is because in this case $p-\Delta > 1/2$ and therefore the Taylor series for $G_I(p,\Delta)$ can be used up to order Δ^4 to obtain the approximate analytical solution. In Fig. 4.2 one can notice that the higher is the winning probability p, the harder it is for the insider to outperform the outsider.

4.2.3 Finite memory problem

As it has been already mentioned, in real life investors lack information on the exact value of the winning probability p—it has to be inferred from the available past data. In addition, since p can vary in time, it may be better to focus on a recent part of the data and obtain a fresh estimate. To model the described situation we assume that the investor uses only game outcomes from the last L turns for the inference and that the winning probability p is fixed during this period (a generalization to variable p will be also discussed). The impact of uncertainty on the Kelly portfolio is closely investigated in [145] where certain prior information and long-term stationarity of p are assumed.

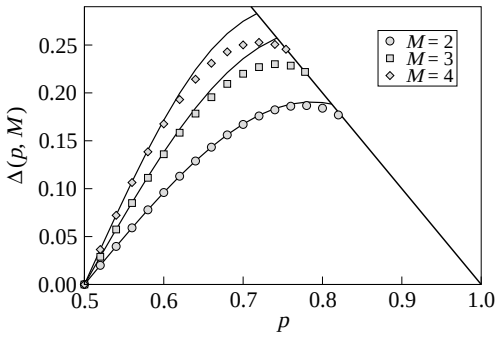


Figure 4.2: The limiting value $\Delta(p, M)$ when the investment performances of the insider investing in one game and the outsider investing in M games are equal. Numerical solutions of Eq. (4.15) are shown as symbols, analytical results obtained as described in the text are shown as lines.

Let's label the number of winning games in last L turns as w ($w = 0, \ldots, L$). The resulting information about p can be quantified using the Bayes theorem (for a discussion of the Bayes theorem see *e.g.*[146]) as

$$\varrho(p|w, L) = \frac{\pi(p)P(w|p, L)}{\int_0^1 \pi(p)P(w|L, p)\,\mathrm{d}p}. \tag{4.16}$$

Here $\pi(p)$ is the prior probability distribution of p and $P(w|L, p)$ is the probability distribution of w, given the values p and L. Due to mutual independence of consecutive outcomes, $P(w|L, p)$ is a binomial distribution and $P(w|L, p) = \binom{L}{w} p^w (1-p)^{L-w}$. All information available to the investor is represented by the observation of w winning games in last L turns—there is no additional information entering the inference. Therefore we assume the maximum prior ignorance by choosing $\pi(p) = 1$ for $p \in [0; 1]$ (a so-called uniform prior). For the chosen distributions $P(w|p, L)$ and $\pi(p)$, Eq. (4.16) simplifies to the form

$$\varrho(p|w, L) = \frac{(L+1)!}{w!(L-w)!}\, p^w (1-p)^{L-w}. \tag{4.17}$$

This represents the gambler's information about p after observing w wins in the last L turns.

Now instead of one fixed value for the winning probability p, only the probability distribution $\varrho(p)$ is known. In such a case, maximization of $G = \langle \ln W_1 \rangle$ results in $f^* = 2\langle p \rangle - 1$. We prove this theorem for a special case of two possible winning probabilities p_1 and p_2: $P(p_1) = P_1$, $P(p_2) = P_2$, $P_1 + P_2 = 1$ (extension to the general case is straightforward). The exponential growth rate can be now written as

$$G = (P_1 p_1 + P_2 p_2)\, \ln(1+f) + (1 - P_1 p_1 - P_2 p_2)\, \ln(1-f).$$

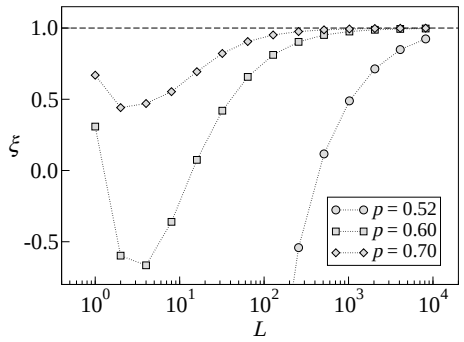


Figure 4.3: The ratio $\xi := \mathcal{R}(p, L)/\mathcal{R}_{\mathrm{K}}(p)$ as a function of the memory length L.

This is maximized by $f^* = 2(P_1 p_1 + P_2 p_2) - 1$. Since $P_1 p_1 + P_2 p_2 = \langle p \rangle$, we have proven that

$$f^* = 2\langle p \rangle - 1. \tag{4.18}$$

From Eq. (4.17) follows $\langle p \rangle = (w+1)/(L+2)$ and consequently

$$f^*(w, L) = \frac{2w - L}{L + 2} \tag{4.19}$$

for $w \geq L/2$. Since we do not consider the possibility of short selling ($f < 0$) in this work, $f^* = 0$ for $2w < L$. Notice that even when $w = L$ (all observed game are winning), $f^* < 1$. This is a direct consequence of the gambler's noisy information about p.

It is instructive to compute the exponential growth rate $G(p, L)$ of an investor with the memory length L. If the winning probability p is fixed during the game, we have

$$G(p, L) = \sum_{w=0}^{L} P(w|p, L) \left[p \ln(1 + f^*(w, L)) + (1 - p) \ln(1 - f^*(w, L)) \right]. \tag{4.20}$$

Consequently, the compounded return is $\mathcal{R}(p, L) = \exp[G(p, L)] - 1$. This result can be compared with $\mathcal{R}_{\mathrm{K}}(p)$ of an investor with the perfect knowledge of p given by Eq. (4.8). In Fig. 4.3a, the ratio $\xi := \mathcal{R}(p, L)/\mathcal{R}_{\mathrm{K}}(p)$ is shown as a function of L for various p. Notice that when p is smaller than a certain threshold (which we numerically found to be approximately 0.63), $\mathcal{R}(p, L) < 0$ for some L. As L increases, the investor's information about p improves and $\xi \to 1$. It can be shown (see [147]) that $G(p, L) \approx G_{\mathrm{K}}(p, L) - 1/(2L)$.

When p is small, a very long memory is needed to make a profitable investment: *e.g.*for $p = 0.51$, $L \geq 1\,761$ and for $p = 0.52$, $L \geq 438$ is needed. This is in agreement with the experience of finance practitioners, according to them the Kelly portfolio is sensitive to a wrong examination of the investment profitability (a scientific analysis of this problem can be found in [148]). However, one

should not forget about the a priori distribution $\pi(p)$ which is an efficient tool to control the investment. For example, to avoid big losses in a weakly profitable game, $\pi(p)$ constrained to the range $[0; 3/4]$ can be used. In turn, if the game happens to be highly profitable and $p > 3/4$, such a choice of $\pi(p)$ reduces the profit.

Another interpretation of the finite memory problem

The optimal investment of a gambler with the memory length L can be inferred also by the direct maximization of the exponential growth rate. In addition to Eq. (4.20), from the investor's point of view G needs to be averaged over all possible values of p, leading to

$$G(L) = \int_0^1 \pi(p)\,\mathrm{d}p \sum_{w=0}^{L} P(w;p,L)\left(p\ln[1+f(w,L)]+(1-p)\ln[1-f(w,L)]\right). \quad (4.21)$$

This quantity can be maximized with respect to the investment fractions $f(w, L)$ which is equivalent to the set of equations $\partial G(L)/\partial f(w, L) = 0$ ($w = 0, \ldots, L$). For $\pi(p) = 1$ in the range $[0; 1]$ this set can be solved analytically and yields the same optimal investment fractions as given in Eq. (4.19).

The statistical models described above use $\pi(p)$ as a model for the gambler's a priori knowledge of the winning probability p. This a priori knowledge can be caused by the lack of gambler's information but also it can stem from the fact that p changes in time. Then $\pi(p)$ represents the probability that at a given moment, the winning probability is equal to p. Since such an evolution of game properties is likely to occur in real life, we investigate it in detail in the following paragraph. Notice that the possible changes of p in time are the key reason why a gambler should use only a limited recent history of the game.

As explained above, the evolution of p can be well incorporated to $\pi(p)$. Consequently, if the changes of p are slow enough to assume that within time window of the length L the winning probability is approximately constant, all the analytical results above hold in the same form and thus the optimal investment is given by Eq. (4.19). To test this conclusion, we maximized G numerically with $f(w, L)$ as variables ($w = 0, \ldots, L$) for five separate realizations of the game, each with the length 1 000 000 turns and $L = 10$. In each realization, the winning probability changed regularly and followed the succession $0.5 \to 1 \to 0 \to 0.5$. As a maximization method we used simulated annealing [149, 150]. In Fig. 4.4, the result is shown together with $f^*(w, L)$ given by Eq. (4.19) and a good agreement can be seen. Thus we can conclude that with a proper choice of $\pi(p)$, the analysed model describes also a risky game with a slowly changing winning probability.

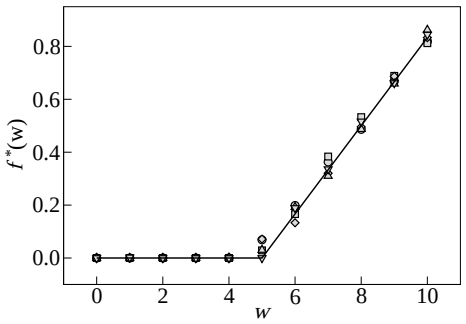


Figure 4.4: A comparison of the analytical optimal fractions $f^*(w, L)$ with results of numerical maximization of G for $L = 10$ and $N = 1\,000\,000$. Various symbols are shown for five independent realizations of the risky game.

4.2.4 Conclusion

In this work we examined maximization of the exponential growth rate, originally proposed by Kelly, in various scenarios. Our main goal was to explore the effects of diversification and information on the investment performance. To ease the computation, instead of working with real assets we investigated simple risky games with binary outcomes: win or loss. This allowed us to obtain analytical results in various model situations.

In the case when multiple independent investment opportunities are simultaneously available we proposed two complementary approximations which yield analytical results for the optimal investment fractions. Based on these results, we proposed a simple framework which can be used to investigate the competition of the uninformed investor (the outsider) who diversifies his portfolio and the informed investor (the insider) who focuses on one investment opportunity. We found the conditions when gains from the diversification exceed gains from the additional information and thus the outsider outperforms the insider.

Finally we investigated the performance of the Kelly strategy when the return distribution (in our case the winning probability of a risky game) is not known a priori. When the past game outcomes represent the only source of information, we found a simple analytical formula for the optimal investment. We showed that for a weakly profitable game, a very long history is needed to allow a profitable investment. As game properties may change in time and thus the estimates obtained using long histories may be biased, this is an important limitation. With short period estimates suffering from uncertainty and long period estimates suffering from non-stationarity, the Kelly strategy may be unable to yield a profitable investment. Closer investigation of the Kelly strategy under realistic market conditions remains a future challenge.

4.3 Kelly-optimal portfolios for lognormally distributed returns

Sometimes your best investments are the ones you don't make.
Donald Trump

In this section we discuss outcomes of the Kelly strategy in the case when investment returns are lognormally distributed. We derive analytical results for optimal portfolios in various situations, as well as numerical solutions and computer simulations. In addition we show that, in the limit of small returns and volatilities, when borrowing is not allowed, the Kelly-optimal portfolio lies on the Efficient Frontier. Furthermore, we analyse the conditions under which diversification is no longer profitable and the optimal portfolio condensates on a few assets.

4.3.1 Simple model

We shall study the portfolio optimisation on a very simple model which leads to lognormally distributed returns. Consider N assets, whose prices $p_i(t)$ ($i = 1, \ldots, N$) undergo uncorrelated geometric Brownian motions

$$p_i(t) = p_i(t-1)\,\mathrm{e}^{\eta_i(t)}. \tag{4.22}$$

Here the random numbers $\eta_i(t)$ are drawn from Gaussian density distributions of mean m_i and variance D_i, and are independent of their value at previous time steps. This model can be easily generalised to the case of non-Gaussian densities and correlated price variations as it is discussed in Sec. 4.5.2. We further assume the existence of a risk-free asset paying zero interest rate.

For the sake of simplicity, we do not include dividends, transaction costs and taxes in the model. Hence, the return of asset i is

$$R_i(t) := \frac{p_i(t) - p_i(t-1)}{p_i(t-1)} = \mathrm{e}^{\eta_i(t)} - 1.$$

It follows that the average return of this asset is $\mu_i := \langle R_i \rangle = \exp[m_i + D_i/2] - 1$ and the volatility is $\sigma_i^2 := \langle (R_i - \mu_i)^2 \rangle = (\exp[D_i] - 1)\exp[2m_i + D_i]$. Angular brackets denote averages over the noise $\eta_i(t)$.

A portfolio is determined by the fractions f_i of the total capital invested in each one of N available assets; the rest is kept in the risk-free asset. As both Kelly strategy and the Efficient Frontier use one time step optimisation only, for us the basic quantity is the wealth after one time step W_1. If we set the initial wealth to 1, W_1 has the form

$$W_1 = 1 + \sum_{i=1}^{N} f_i R_i = 1 + R_P, \tag{4.23}$$

where $R_P := \sum_{i=1}^{N} f_i R_i$ is the portfolio return. To simplify the computation we assume infinite divisibility of the investment. Thus, the investment fractions f_i are real numbers and do not need to be rounded.

In the portfolio optimisation, some common constraints are often imposed and can as well be applied in the present context. For instance, non-negativity of the investment fractions $f_i \geq 0$ forbids short positions. The condition $\sum_{i=1}^{N} f_i = 1$ indicates the absence of a riskless asset and $\sum_{i=1}^{N} f_i \leq 1$ does not allow the investor to borrow money.

4.3.2 The Mean-Variance approach

The unconstrained maximisation of the expected capital gain results in the investment of the entire wealth on the asset with the highest expected return; this strategy is sometimes referred to as risk neutral. If the investor has a strong aversion to risk, on the other hand, one might be tempted to simply minimise the portfolio variance $\sigma_P^2 = \sum_{i=0}^{M} f_i^2 \sigma_i^2$. This leads to invest the entire capital on the risk-free asset with no chance to benefit from asset price movements. The Mean-Variance (MV) approach is much more reasonable as it allows to compromise between the gain and the risk. Here we remind basic results of this standard tool.

With the desired expected return fixed at $\langle R_P \rangle = \mu_P$, the constrained minimisation of the portfolio variance σ_P^2 is performed using the Lagrange function $\mathcal{L} = \langle R_P^2 \rangle + \gamma\big(\langle R_P \rangle - \mu_P\big)$ with a Lagrange multiplier γ. The resulting optimal fractions are

$$f_i^* = \mu_P \frac{\mu_i}{C_2 \sigma_i^2}, \quad \text{where } C_k = \sum_{j=1}^{N} \frac{\mu_j^k}{\sigma_j^2}. \tag{4.24}$$

For $\mu_P = 0$, $f_i^* = 0$ for all assets. As we increase μ_P, all optimal fractions f_i^* grow in a uniform way and their ratios are preserved. At some value μ_P^* we reach $\sum_{i=1}^{N} f_i^* = 1$, which means we are investing the entire capital. Any further increase would require to borrow money, with Eq. (4.24) remaining valid as long as the borrowing rate equals the lending rate (both set to zero here). The relation between σ_P and μ_P is

$$\sigma_P = \mu_P / \sqrt{C_2}. \tag{4.25}$$

This equation is often referred to as Capital Market Line (CML).

If there is no risk-free asset in the market, one has to introduce the additional constraint $\sum_{i=1}^{N} f_i^* = 1$. It follows that

$$\sigma_P^2 = \frac{C_0 \mu_P^2 - 2C_1 \mu_P + C_2}{C_0 C_2 - C_1^2}. \tag{4.26}$$

The functional relation between the optimised σ_P and μ_P is called Efficient Frontier (EF). Since there is only one point on the CML where $\sum_i f_i = 1$, this line

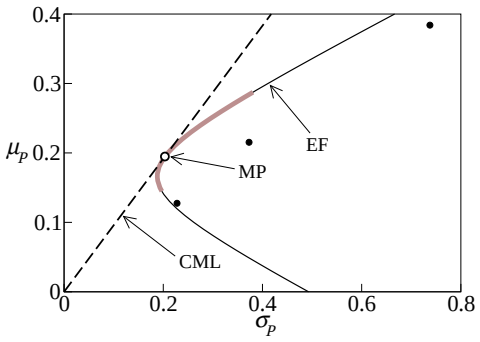


Figure 4.5: The expected return μ_P versus the standard deviation of the portfolio σ_P. The assets are described by the following parameters: $m_1 = 0.1$, $D_1 = 0.04$, $m_2 = 0.15$, $D_2 = 0.09$, $m_3 = 0.2$, $D_3 = 0.25$ (in the graph they are shown as full circles). The dashed line represents the CML from Eq. (4.25), the solid line is Efficient Frontier given by Eq. (4.26), the tangent point of the two is the Market Portfolio. The thick part of EF marks the region where all investment fractions are positive.

is tangent to the EF. According to ref. [125], investors should only choose this point, the Market Portfolio (MP), as the optimal portfolio of risky assets, with the reminder of their wealth kept in cash. The results of this section are plotted in Fig. 4.5 for a particular choice of three available assets.

4.3.3 The Kelly portfolio

Since the expected value of the investor's capital is dominated by rare events, it is not reasonable to form a portfolio by simply maximising $\langle W(t) \rangle$. The Mean-Variance approach tries to solve this problem in a straightforward, yet criticisable way. We support here the idea that an efficient investment strategy can be found by maximising the expected wealth cumulated in the long run, which is, according to Kelly [130], equivalent to maximising the logarithm of the wealth W_1 after one time step. Thus the key quantity in the construction of a Kelly-optimal portfolio is $v := \langle \ln W_1 \rangle$, the average exponential growth rate of the wealth. We remind that the quantity $\ln W_1$ is not a logarithmic utility function.

We should spend a word of caution about v. When $\sum_{i=1}^{N} f_i$ lies out of the range $[0, 1]$, there is a nonzero probability that W_1 is negative and, as a consequence, its logarithm is not defined. In order to obtain a generalisation of the presented results to the case with $f_i \notin [0, 1]$ one needs to cope with the influence of this singularity.

The authors of ref. [140] optimise v in a similar context, claiming that their procedure corresponds to maximising the median of the distribution of returns.

They consider short time intervals and thus small assets returns. Since $R_P \ll 1$, they use the approximation $\ln(1 + R_P) \approx R_P - R_P^2/2$ of the logarithm in the expression of v before maximising it. However, while such an expansion is only justified for $R_P \ll 1$, the maximum of the resulting function is at $R_P = 1$, in contradiction with the hypothesis. We will develop a different approximation in the following.

First, the unconstrained maximisation of v is achieved by solving the set of equations $\partial v/\partial f_i = 0$ ($i = 1, \ldots, N$). After exchanging the order of the derivative and the average, we obtain the condition

$$\left\langle \frac{R_i}{1 + \sum_j f_j R_j} \right\rangle = 0 \qquad (i = 1, \ldots, N). \tag{4.27}$$

In our case, R_i has a lognormal distribution and this set of equations can not be solved analytically. With the help of the approximations introduced in Sec. 4.5.1 we shall work out approximative solution for some particular cases.

One risky asset

Let us begin the reasoning with the case of one risky asset. We want to find the optimal investment fraction f of the available wealth. The remaining fraction $1-f$ we keep in cash at the risk-free interest rate which, without loss of generality, is set to zero. This problem is described by Eq. (4.27) in one dimension; even this simplest case has no analytical solution. Nevertheless, given the asset properties m and D, one can ask what is the minimal value $m_<$ for which it becomes profitable to invest a positive fraction of the investor's capital in the risky asset. This can be found imposing $f = 0$ in Eq. (4.27), yielding $m_< = -D/2$. The opposite limit corresponds to investing the entire capital in the risky asset; imposing $f = 1$ yields $m_> = D/2$.

We shall look for approximate solutions that are valid for small values of D, which is the case treated in Sec. 4.5.1. Using approximation Eq. (4.63) in Eq. (4.27) gives

$$\frac{\mathrm{e}^m - 1}{1 - f + f\mathrm{e}^m} + \frac{D}{2} \frac{\mathrm{e}^m(1 - f - f\mathrm{e}^m)}{(1 - f + f\mathrm{e}^m)^3} = 0.$$

With respect to f, this is merely a quadratic equation. Since the solution is rather long, we first simplify the equation using $m, D \ll 1$ as in Eq. (4.64), leading to the result

$$f^* = \frac{1}{2} + \frac{m}{D}. \tag{4.28}$$

When the assumption of asset prices undergoing a geometric Brownian motion is valid, both m and D scale linearly with the time scale and f^* does not depend on the length of the time step. Notice also that substituting $m = \pm D/2$ gives $f^* = 0$ and $f^* = 1$, in agreement with the bounds we found before by exact

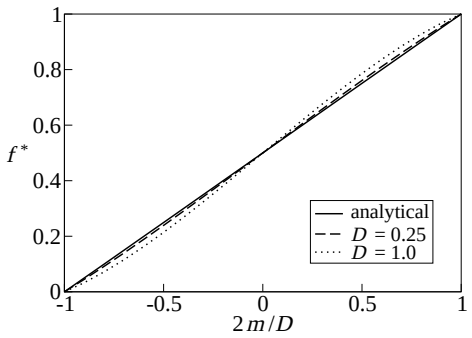


Figure 4.6: The optimal portfolio fraction f^*: a comparison of the analytical result Eq. (4.28) with a numerical maximisation of $\langle \ln W_1 \rangle$.

computation. The first order correction to Eq. (4.28) is $m(4m^2 - D^2)/4D^2$ which is, for $m \in [-D/2, D/2]$, of order $O(m)$. The validity of the presented approximations can be easily tested by a straightforward numerical maximisation of $\langle \ln W_1 \rangle$. As can be seen in Fig. 4.6, the numerical results are well approximated by the analytical formula Eq. (4.28) even for $D = 1$.

Comparison with known results

First of all, for $m, D \ll 1$, one can approximate $\mu \approx m + D/2$ and $\sigma^2 \approx D$, which makes the optimal portfolio fraction derived above equal to the one obtained in [140]: $f' = \mu/(\mu^2 + \sigma^2)$. However, if we check the accuracy of f', we find a relative error up to 3% for $D = 0.01$, and for $D = 0.25$ we are already far out of the applicability range with an error around 50%. Also, Eq. (4.28) is for $m, D \ll 1$ identical to the well-known Merton's result [151] which is derived using the assumption of a logarithmic utility function.

Investment with continuous readjustment is described by the stochastic differential equation

$$\mathrm{d}W(t) = fW(t)\left(\mu_I \,\mathrm{d}t + \sigma_I \,\mathrm{d}Z\right) \tag{4.29}$$

where f is the investment fraction, $\mathrm{d}Z$ is the standard white noise, and the equation is written in the Itô sense [152, 153]. Since readjustment is continuous, maximization of $\langle \ln W(T) \rangle$ for a fixed deadline T is similar to maximization of $\langle \ln W_N \rangle$ in the limit $N \to \infty$. Using the convergence of random walk to white noise, in [145] it is shown that the optimal investment fraction is equal to $f^* = \mu_I/\sigma_I^2$. To make wealth governed by Eq. (4.29) during the time period Δt equivalent to the discrete time evolution described by Eq. (4.22), one must set $\mu_I \,\Delta t = m + D/2$ and $\sigma_I^2 \,\Delta t = D$. Consequently, the exact optimal fraction derived in [145] is the same as our approximate result given in Eq. (4.28). This is not surprising because continuous readjustment is equivalent to Eq. (4.22) with $m, D \to 0$.

Constrained optimisation

The optimal portfolio fractions f_i can be derived from Eq. (4.27) also for $N > 1$. Using the same approximations as in the single asset case, we obtain the general formula $f_i^* = 1/2 + m_i/D_i$ for $i = 1, \ldots, N$. When $\sum_{i=1}^N f_i > 1$, one has to borrow money to follow the prescribed strategy. If borrowing is not allowed, we have to introduce the constraint $\sum_{i=1}^N f_i = 1$. This can be done by use of the Lagrange function $\mathcal{L}(\boldsymbol{f}, \gamma) = v + \gamma \left(\sum_{i=1}^N f_i - 1 \right)$. The optimal portfolio is then the solution of the set of equations

$$\sum_{j=1}^{N} f_j = 1, \quad \left\langle \frac{R_i}{1 + \sum_{i=1}^N f_i R_i} \right\rangle + \gamma = 0 \qquad (i = 1, \ldots, N), \tag{4.30}$$

where $R_i = \mathrm{e}^{\eta_i} - 1$. Using the same approximations one obtains

$$f_i^* = \frac{1}{2} + \frac{m_i + \gamma}{D_i}. \tag{4.31}$$

The Lagrange multiplier γ is fixed by the condition $\sum_j f_j^* = 1$. It can occur that even a profitable asset with $m_i > -D_i/2$ has a negative optimal investment fraction: a short position is suggested. If short selling is not allowed, this asset has to be eliminated from the optimisation process. In consequence, under some conditions, only a few assets are included in the optimal portfolio. This phenomenon, which we call portfolio condensation, we study closer in sections 4.3.3 and 4.3.3.

Now we can establish an important link to Markowitz's approach: in the limit $\mu_i, \sigma_i \to 0$ the constrained Kelly portfolio lies on the Efficient Frontier. We shall prove this statement in the following. When all the assets have small μ_i and σ_i, in Eq. (4.24) and Eq. (4.26) we can approximate $\mu_i \approx m_i + D_i/2$ and $\sigma_i^2 \approx D_i$, leading to the approximative relation for the Efficient Frontier

$$\sigma_P^2 = \frac{\tilde{C}_0 \mu_P^2 - 2\tilde{C}_1 \mu_P + \tilde{C}_2}{\tilde{C}_0 \tilde{C}_2 - \tilde{C}_1^2}, \quad \text{where } \tilde{C}_k = \sum_{j=1}^{N} \frac{(m_i + D_i/2)^k}{D_i}. \tag{4.32}$$

For the Kelly portfolio we need to work out a similar approximation. Using the condition $\sum_i f_i^* = 1$, for γ in Eq. (4.31) we obtain $\gamma = (\tilde{C}_1 - 1)/\tilde{C}_0$. In the relations $\mu_P = \sum_i f_i \mu_i$ and $\sigma_P^2 = \sum_i f_i^2 \sigma_i^2$ we use the approximations for μ_i, σ_i introduced above. After substituting f_i from Eq. (4.31), for the Kelly optimal portfolio we get

$$\mu_K = \frac{\tilde{C}_0 \tilde{C}_2 - \tilde{C}_1^2 + \tilde{C}_1}{\tilde{C}_0}, \quad \sigma_K^2 = \frac{\tilde{C}_0 \tilde{C}_2 - \tilde{C}_1^2 + 1}{\tilde{C}_0}. \tag{4.33}$$

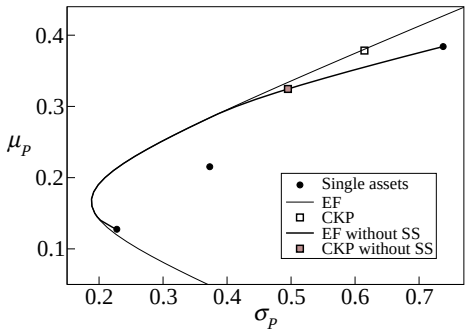


Figure 4.7: The Efficient Frontier (EF) and the constrained Kelly portfolio (CKP) in a particular case of three assets (asset parameters as in Fig. 4.5). The thin solid line represents the EF. When short selling is forbidden, we obtain a modified EF shown by the bold line. The open square is the constrained Kelly optimal portfolio, while the filled one is the Kelly's optimum when short selling is forbidden.

Now it is only a question of simple algebra to show that μ_K and σ_K given by Eq. (4.33) fulfill Eq. (4.32), which completes the proof. Similar, yet weaker, results can be found in the literature. For instance, Markowitz states in [138] that "On the EF there is a point which approximately maximizes $\langle \ln W_1 \rangle$."

The results obtained so far are illustrated in Fig. 4.7, where we plot the Efficient Frontier together with the constrained Kelly portfolio for the same three assets as in Fig. 4.5. We show also results with one additional constraint: no short selling allowed (*i.e.*, investment fractions are non-negative). While the original EF is not bounded (for any μ_P we can find an appropriate σ_P), the EF with the additional constraint starts at the point corresponding to the full investment in the least profitable asset and ends at the point corresponding to the most profitable asset. The two lines coincide on a wide range of μ_P. In agreement with the previous paragraph, the constrained Kelly portfolio lies close to the EF and the same holds in the case with the additional constraint. For illustration, the unconstrained Kelly portfolio for the same three assets has $\mu_P \approx 1.35$ and $\sigma_P \approx 1.43$ with the total investment fraction $\sum_i f_i^* \approx 6.5$ far exceeding the actual capital.

Condensation in the two asset case

To illustrate the condensation phenomenon we focus on a simple case here: two risky assets plus a risk-free one, borrowing and short selling forbidden. As we have already seen, without constraints $f_i^* = 1/2 + m_i/D_i$. Therefore, when $m_i < -D_i/2$, f_i is negative and due to forbidden short selling, asset i drops out of the optimal portfolio. In Fig. Eq. (4.8) this threshold is shown for $i = 1, 2$ by dashed lines. In the lower-left corner (A) we have the region where both assets

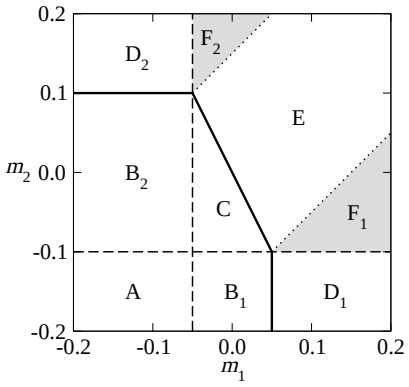


Figure 4.8: The phase diagram of the two-asset system with $D_1 = 0.1$ and $D_2 = 0.2$. In region A the investor is advised to use only the risk-free asset. In regions B_1, B_2, and C the optimal investment is still partially in the risk-free asset. In regions D_1, D_2, E, F_1, and F_2 one should invest everything in the risky assets. While in regions C and E the investment is divided between the two assets, in shaded regions F_1 and F_2 a nontrivial condensation arises: one is advised to invest all wealth in one asset although the other one is also profitable.

are unprofitable and the optimal strategy prescribes a fully riskless investment.

When the results of the unconstrained optimisation sum up to one ($f_1^* + f_2^* = 1$), we are advised to invest all our wealth in the risky assets. If both assets are profitable, this occurs when $m_1/D_1 + m_2/D_2 = 0$. When only asset i is profitable, we should invest the entire capital on it only when m_i equals, at least, $D_i/2$. In Fig. Eq. (4.8) these results are shown as a thick solid curve.

Since borrowing is not allowed, in the region above the solid line constrained optimisation has to be used. The condensation to one of the two assets arises when the optimal fractions (f_1^*, f_2^*) are either $(1, 0)$ or $(0, 1)$. We can find the values m_1' and m_1'' when this happens. By eliminating γ from Eq. (4.30) and substituting $f_1 = 1$ and $f_2 = 0$ we obtain the condition for the condensation on asset 1: $\left\langle (\mathrm{e}^{\eta_1} - 1)/\mathrm{e}^{\eta_1} \right\rangle = \left\langle (\mathrm{e}^{\eta_2} - 1)/\mathrm{e}^{\eta_1} \right\rangle$. This can be solved analytically, yielding

$$m_1' = m_2 + \frac{D_1 + D_2}{2}. \tag{4.34}$$

This equation holds with interchanged indices for the condensation on asset 2, thus $m_1'' = m_2 - (D_1 + D_2)/2$. Finally, for $m_1'' < m_1 < m_1'$ the optimal portfolio contains both assets. The crossover values m_1' and m_2' are shown in Fig. 4.8 as dotted lines. They delimit the region where the portfolio condensates to only one of two profitable assets. A complete "phase diagram" of the optimal investment in the two assets case is presented in Fig. 4.8 for a particular choice of the assets' variances.

Many assets with equal volatility

We investigate here the case of an arbitrary large number N of available assets. We assume that borrowing and short selling are forbidden, which makes it possible to see a portfolio condensation. While the optimal portfolio fractions are given by Eq. (4.31), to find which assets are included in the optimal portfolio is a hard combinatorial task. To obtain analytical results, we simplify the problem with the assumption that the variances of all assets are equal, $D_i = D$ ($i = 1, \ldots, N$). The number of assets contained in the optimal portfolio is labelled as M.

Let us first sort the assets in order of decreasing m_i, such that $m_1 > m_2 > \cdots > m_N$. We form the optimal portfolio starting from the most profitable asset m_1, and adding the others one by one until the last added asset has a negative optimal fraction $f_{M+1} \leq 0$. Summing Eq. (4.31) from 1 to M, we can write γ as $\gamma(M) = D(M^{-1} - 1/2) - \frac{1}{M}\sum_1^M m_i$. For the given realisation of $\{m_i\}$, we can find M by solving the equation $f_M = 0$, which reads

$$m_M + \frac{D}{M} = \frac{1}{M}\sum_{i=1}^{M} m_i. \tag{4.35}$$

This relation tells us how many assets we should invest on, once their expected growths and volatility are known. Notice that for $M = 2$ and $D_1 = D_2 = D$, this result is consistent with that of Eq. (4.34) where a special case of the condensation on two assets is described.

Let us follow now a statistical approach. If all m_i are drawn from a given distribution $\varrho(m)$, the value of M depends on the current realisation. The typical behaviour of the system can be found by taking the average over all possible realizations and replacing m_i by $\overline{m}_i$. The resulting portfolio size M_T captures the typical behaviour and depends on the distribution $\varrho(m)$ and on the number of available assets N.

Uniform distribution of m

Let us first analyse the case of a uniform distribution of m_i within the range $[a, b]$. First we assume that all assets are profitable, *i.e.*$a + D/2 > 0$. For m_i are sorted in decreasing order, one can show $\overline{m}_i = b - (b-a)i/(n+1)$. Since $\overline{m}_i$ grows with i linearly, according to Eq. (4.31), $\overline{f_i^*}$ decreases with i uniformly. Substituting $\overline{m}_i$ for m_i in Eq. (4.35), we obtain an equation for M_T. Assuming $M_T \gg 1$, the solution has the simple form

$$M_T = \sqrt{2ND/(b-a)}. \tag{4.36}$$

As it is shown in Fig. 4.9, this results agrees well with a numerical treatment of the problem (based on Eq. (4.35)).

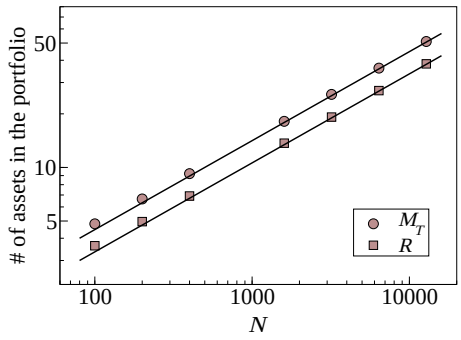


Figure 4.9: The average size of the optimal portfolio M_T and the inverse participation ratio $\mathcal{R}$ for N assets with $D = 0.01$ and m_i uniformly distributed in the range $[0, 0.1]$. Solid lines are the analytical solutions reported in Eqs. Eq. (4.36) and Eq. (4.37), symbols stand for numerical results.

A more flexible measure of the level of condensation is the *inverse participation ratio* (sometimes called the Herfindahl index), defined as $\mathcal{R} = 1/\sum_{i=1}^{N} f_i^2$. It estimates the effective number of assets in the portfolio: when all investment fractions are equal, $\mathcal{R} = M$, while when one asset covers 99% of the portfolio, $\mathcal{R} \approx 1$. Using Eq. (4.31) we can write $f_i = A - Bi$ ($B = (b-a)/(n+1)$, the detailed form of A is not needed for the solution). Passing from $i = 1$ to $i = M$, f_i decreases linearly to zero. Therefore we can use the identity $\sum_{i=1}^{M}(A - Bi)^2 = \sum_{i=1}^{M}(Bi)^2$ and obtain

$$\mathcal{R} = \left[B^2 M(M+1)(2M+1)/6\right]^{-1} \approx \left[B^2 M^3/3\right]^{-1} \approx \tfrac{3}{4}M. \tag{4.37}$$

In the last step we used formula Eq. (4.36) for the typical size of the optimal portfolio. We see that a uniform distribution of m_i leads to the participation ratio proportional to the number of assets in the portfolio. In Fig. 4.9, Eq. (4.37) is shown to match the numerical solution.

Power-law distribution of m

Now we treat the case of a distribution $\varrho(m)$ that has a power-law tail: $\varrho(m) = Cm^{-\alpha-1}$ for $m > m_{\min}$. As long as $M \ll N$, the properties of the assets included in the optimal portfolio are driven by the tail of $\varrho(m)$. In consequence, the detailed form of $\varrho(m)$ for $m < m_{\min}$ is not important here. We assume that only a fraction r of all assets falls in the region $m > m_{\min}$.

Instead of seeking the typical portfolio size M_T, we shall limit ourselves to finding the conditions when a condensation on one asset arises. With this aim in mind, we put $M = 2$ in Eq. (4.35), obtaining the equation $m_1 - m_2 = D$. When $m_1 - m_2 > D$, only asset 1 is included in the optimal portfolio. By replacing m_1 and m_2 with their medians $\tilde{m}_1$ and $\tilde{m}_2$, one obtains an approximate condition for

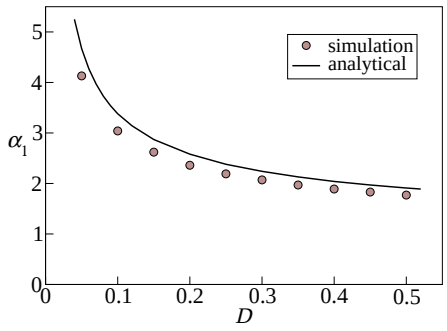


Figure 4.10: Values of the power law exponent α_1 at which condensation to one asset arises for $N = 1\,000$, $r = 0.1$, and $m_{\min} = 0.1$. We compare the analytical result (shown as the solid line) with a numerical simulation of the system (shown as filled circles). Below the line, the optimal portfolio typically contains only one asset.

a system where such a condensation typically exists. Using *order statistics* [154] we find the following expressions for the medians: $\tilde{m}_1 = m_{\min}(Nr/\ln 2)^{1/\alpha}$, $\tilde{m}_2 \doteq m_{\min}(Nr/1.68)^{1/\alpha}$. The equation $\tilde{m}_1 - \tilde{m}_2 = D$ thus achieved can be solved numerically with respect to α. In this way we find the value α_1 below which the optimal portfolio typically contains only the most profitable asset. In Fig. 4.10 we plot the result as a function of D. For comparison, the outcomes from a purely numerical investigation of the equation $P(m_1 - m_2 > D) = 0.5$ are also shown as filled circles. Our approximate condition has the same qualitative behaviour as the simulation, showing that the use of median gives us a good notion of the optimal portfolio behaviour.

4.3.4 Conclusion

In this section we investigated the Kelly optimisation strategy in the framework of a simple stochastic model for asset prices. We derived highly accurate approximate analytical formula for the optimal portfolio fractions. We proved that in the limit of small returns and volatilities of the assets, the constrained Kelly-optimal portfolio lies on the Efficient Frontier. Based on the obtained analytical results, we proposed a simple algorithm for the construction of the optimal portfolio in the constrained case. We showed that when borrowing and short positions are forbidden, only a part of the available assets is included in the optimal portfolio. In some cases the size of the optimal portfolio is much smaller than the number of available assets –we say that a portfolio condensation arises. In particular, when the distribution of the mean asset returns is wide, there is a high probability that only the most profitable asset is included in the Kelly-optimal portfolio.

4.4 How to quantify the influence of correlations on diversification

Don't put all your eggs in one basket.

Adage

Most portfolio optimization strategies result in diversification because it allows investors to decrease their exposure to the risk of single assets [155]. However, when assets are correlated, the improvement of investment performance due to diversification is reduced. Since asset correlations are ubiquitous, ranging from stock-stock correlations in one stock market to correlations between different investments types in different countries, it is important to investigate their influence on diversified portfolios [156–160]. In this section we attempt to quantify how correlations reduce the benefits of diversification. To achieve this we propose a new quantity, the effective size of a diversified portfolio, which is based on the comparison with a fictitious portfolio of uncorrelated assets. For two different optimization strategies (the Mean-Variance portfolio [124] and the Kelly portfolio [130]) we obtain analytical expressions for their effective sizes. Obtained results are used to study real market data (stocks from the Dow Jones Industrial Average and the S&P 500).

4.4.1 Correlations and the Mean-Variance portfolio

First we introduce the notation used in this section. If the initial and final asset values are w_0 and w_1 respectively, the asset return during the given time period is defined as $R := (w_1 - w_0)/w_0$. If we follow the value of asset i over many subsequent time periods, it is possible to find the expected values $\mu_i := \langle R_i \rangle$ and $V_i := \langle R_i^2 \rangle - \langle R_i \rangle^2$, closely related is the standard deviation $\sigma_i = \sqrt{V_i}$. The standard tool to measure correlations is the Pearson's formula; for assets i and j it reads

$$C_{ij} := \frac{\langle R_i R_j \rangle - \langle R_i \rangle \langle R_j \rangle}{\sigma_i \sigma_j}, \tag{4.38}$$

by definition $C_{ij} \in [-1; 1]$ and $C_{ii} = 1$. These values form the correlation matrix C. When returns R_i and R_j are independent, $\langle R_i R_j \rangle = \langle R_i \rangle \langle R_j \rangle$ and $C_{ij} = 0$. The same holds when one of the assets is risk-free, *i.e.*its return has zero variance. Although in practice each investment carries a certain amount of risk, short term government-issued securities are often used as proxies for risk-free assets. Here we assume that a risk free asset is available and has zero return.

To construct a portfolio, the investor has to divide the current wealth W among M available assets. This division can be characterized by investment fractions: f_i is the fraction of wealth invested in asset i ($i = 1, \ldots, M$). Assum-

ing unit initial wealth, after one time period it becomes

$$W_1 = 1 + \sum_{i=1}^{M} f_i R_i. \tag{4.39}$$

Here R_i is the return of asset i during the period. When investment fractions are fixed, the investor's wealth follows a multiplicative random walk and after T periods it becomes

$$W_T = \prod_{t=1}^{T} \Big(1 + \sum_{i=1}^{M} f_i R_{i,t}\Big). \tag{4.40}$$

Here $R_{i,t}$ is the return of asset i in the period t.

The Mean-Variance approach to the portfolio optimization has been proposed in [124] (for later discussions see [129, 158, 161]). Despite its flaws (*e.g.*, only the expected return and its variance are used to characterize a portfolio) it is still a benchmark for other optimization methods. From Eq. (4.39), the expected return $R_P := \langle W_1 \rangle - 1$ and the variance $V_P := \langle W_1^2 \rangle - \langle W_1 \rangle^2$ follow as

$$R_P = \sum_{i=1}^{M} f_i \mu_i, \quad V_P = \sum_{i,j=1}^{M} f_i f_j C_{ij} \sigma_i \sigma_j. \tag{4.41}$$

For a given R_P, the optimal portfolio is defined as the one which minimizes V_P (equivalently, one can maximize R_P with V_P fixed). To focus purely on the influence of correlations we assume that all M games are identical ($\mu_i = \mu$ and $\sigma_i = \sigma$). Using Lagrange multipliers, for a desired expected return R_P the minimal portfolio variance is

$$V_P^*(R_P, M, \mathsf{C}) = \frac{\sigma^2 R_P^2}{\mu^2 \sum_{i,j=1}^{M} (\mathsf{C}^{-1})_{ij}} \tag{4.42}$$

where C^{-1} is the inverse of the correlation matrix C.

Now we look at Eq. (4.42) from a new perspective: we compare it with the optimal variance of the portfolio of m uncorrelated assets with identical mean returns and variances $V_P^*(R_P, m, \mathsf{I})$; here I is the identity matrix. The portfolio size when these variances are equal we call the effective size of the correlated portfolio and label it as m_{ef}. By solving the equation $V_P^*(R_p, M, \mathsf{C}) = V_P(R_p, m_{\text{ef}}, \mathsf{I})$ we obtain

$$m_{\text{ef}} = \sum_{i,j=1}^{M} (\mathsf{C}^{-1})_{ij}. \tag{4.43}$$

Otherwise stated, when investing in M correlated assets, the portfolio variance is the same as for m_{ef} uncorrelated assets. Notice that m_{ef} given by Eq. (4.43) depends only on the correlation matrix and not on the portfolio parameters μ, σ, R_P.

Now we review the special case with uniform correlations between the assets: $C_{ij} = C$ for $i \neq j$, $C \in [0; 1]$. Then C^{-1} can be found analytically as

$$(\mathsf{C}^{-1})_{ii} = \frac{1 + (M-2)C}{(1-C)[1 + (M-1)C]}, \quad (\mathsf{C}^{-1})_{ij} = \frac{-C}{(1-C)[1 + (M-1)C]}. \tag{4.44}$$

and Eq. (4.43) consequently simplifies to

$$m_{\text{ef}} = \frac{M}{1 + (M-1)C}. \tag{4.45}$$

When $C = 1$ (perfectly correlated assets), $m_{\text{ef}} = 1$; when $C = 0$ (uncorrelated assets), $m_{\text{ef}} = M$; when $C = -1$ and $M = 2$, the portfolio variance can be totally eliminated and $m_{\text{ef}} \to \infty$. A remarkable consequence of Eq. (4.45) is that in the limit $M \to \infty$, $m_{\text{ef}} \to 1/C$. This means that diversification into arbitrary many assets with mutual correlation C is equivalent to investment in only $1/C$ uncorrelated assets. Another interesting case is a block diagonal matrix C which has square matrices $\mathsf{C}_1, \ldots, C_N$ along the main diagonal and the off-diagonal blocks are zero matrices. It can be shown that Eq. (4.43) then yields

$$m_{\text{ef}} = m_{\text{ef}}(\mathsf{C}_1) + \cdots + m_{\text{ef}}(\mathsf{C}_N). \tag{4.46}$$

Thus, the effective portfolio size is the sum of effective sizes for each block C_i separately. This form of C is an extreme case of the sector structure discussed in Sec. 4.4.2.

It is instructive to compare the results obtained above with the simple investment distributed evenly among all assets. If we again assume identical returns and variances of the assets, Eq. (4.41) simplifies to $R_P = Mf\mu$, $V_P = \sigma^2 f^2 \sum_{i,j=1}^{M} C_{ij}$ where f is the fraction of wealth invested in each single asset. The desired value of R_P now determines both f and V_P. By comparison with the variance of a portfolio of uncorrelated assets we obtain the effective size of the even investment in the form

$$m'_{\text{ef}} = \frac{M}{1 + (M-1)\langle C \rangle} \tag{4.47}$$

where $\langle C \rangle$ is the average of the off-diagonal elements of C, the prime symbol indicates that the even investment is considered.

The effective portfolio size m_{ef} is different from the inverse participation ratio (also called the Herfindahl index) which is defined as $\text{IPR} = 1/\sum_{i=1}^{M} f_i^2$ (f_i is the fraction of wealth invested in asset i). While the former quantifies the influence of correlations, the latter quantifies how unevenly is wealth invested in different assets.

4.4.2 Correlations and the Kelly portfolio

Now we investigate how the concept of the effective portfolio size applies for the Kelly portfolio. As we have already shown, the exponential growth rate G of the investment can be written as $G = \langle \ln W_1 \rangle$ where W_1 is the investor's wealth after one time step. Assuming investment into M simultaneous games, G can be written as

$$G = \left\langle \ln \left[1 + \sum_{i=1}^{M} f_i R_i \right] \right\rangle. \tag{4.48}$$

For $M \geq 5$, one can maximize Eq. (4.48) by numerical techniques [142] or by analytical approximations (see Sec. 4.2.1 for details) as we do below.

To investigate the effect of asset correlations on the Kelly portfolio, we consider M individual assets with the correlation between assets i and j computed by Eq. (4.38) and labeled as C_{ij}. Differentiation of Eq. (4.48) with respect to f_i yields

$$\sum_{\boldsymbol{R}} \frac{P(\boldsymbol{R}) R_i}{1 + \sum_{j=1}^{M} f_j R_j} = 0 \tag{4.49}$$

where $P(\boldsymbol{R})$ is the probability of a given vector of returns $\boldsymbol{R} = (R_1, \ldots, R_M)$ and the summation is over all possible $\boldsymbol{R}$ (when returns are continuous, integration must be used). For $M > 4$, this equation has no analytical solution. Assuming that the investment return $\sum_{i=1}^{M} f_i R_i$ is small (which is plausible if the considered time period is short), we can use the expansion $1/(1+x) \approx 1-x$ to obtain

$$\sum_{j=1}^{M} f_j \langle R_i R_j \rangle = \langle R_i \rangle \qquad (i = 1, \ldots, M). \tag{4.50}$$

This set of linear equations gives the first approximation to f_i: when $\langle R_i \rangle$ and $\langle R_i R_j \rangle$ are known, the optimal investment fractions f_i^* can be obtained. Higher order expansion of $1/(1+x)$ results in higher order cross terms of returns which are generally difficult to compute.

As before, to focus on the influence of correlations we assume identical return distributions of the assets; we set $\langle R_i \rangle = \mu$, $\sigma_i = \sigma$ for $i = 1, \ldots, M$, consequently $\langle R_i R_j \rangle = \mu^2 + \sigma^2 C_{ij}$. After substitution to Eq. (4.49), the optimal investment fractions are

$$\boldsymbol{f}^* = \frac{\mu \, (\mathsf{C}^{-1}\mathbf{1})}{\sigma^2 + \mu^2 \sum_{i,j=1}^{M} (\mathsf{C}^{-1})_{ij}} \tag{4.51}$$

where $\mathbf{1}$ is the M-dimensional vector with all elements equal to 1. The effective portfolio size is obtained m_{ef} by a comparison with a portfolio of uncorrelated assets. One way to do this is by comparing the total invested wealth in both

cases. Thus, m_{ef} is defined as the solution of

$$\sum_{i=1}^{M} f_i^*(M, \mathsf{C}) = m_{\text{ef}}\, f^*(m_{\text{ef}}, \mathsf{I}). \tag{4.52}$$

Using Eq. (4.51) we obtain

$$m_{\text{ef}} = \sum_{i,j=1}^{M} (\mathsf{C}^{-1})_{ij} \tag{4.53}$$

which is exactly the same expression of m_{ef} as Eq. (4.43) in the Mean-Variance approach; the effective portfolio size is a common quantity for these two optimization schemes. However, this exact correspondence is true only when the first order approximation is used to solve Eq. (4.49).

It is possible to define the effective number of assets m_{ef} differently: as the number of mutually uncorrelated assets when the expected exponential growth rate $G(m_{\text{ef}}, \mathsf{I})$ is equal to the growth rate $G(M, \mathsf{C})$ with M correlated assets. These two definitions are similar and yield similar results. For practical reasons (the total investment fraction is easier to handle analytically than the exponential growth rate), we confine our analysis to the former one.

Special case with identical correlations

Assuming identical asset correlations, $C_{ij} = C$ for $i \neq j$ ($C \in [0; 1]$), Eq. (4.53) simplifies to

$$m_{\text{ef}} = \frac{M}{1 + (M-1)C}. \tag{4.54}$$

Contrary to the exact formula Eq. (4.45), this result is based on the first order approximation in Eq. (4.49). To review the accuracy of the resulting m_{ef}, we treat the problem numerically. Since with identical correlation all assets are equal, the optimal investment is distributed evenly among them. Thus, maximization of the exponential growth rate G simplifies to a one-variable problem and Eq. (4.49) is replaced by

$$\sum_{\boldsymbol{R}} \frac{P(\boldsymbol{R}) \sum_{j=1}^{M} R_j}{1 + f \sum_{j=1}^{M} R_j} = 0. \tag{4.55}$$

To proceed, one needs to specify the joint distribution of returns, $P(\boldsymbol{R})$. To do so we use simple assets with binary outcomes: $R_i = 1$ (which we label as $+_i$) and $R_i = -1$ (which we label as $-_i$). To induce the correlations we use an artificial hidden asset h with the outcome $+_h$ with the probability p and $-_h$ with the probability $1 - p$. Finally, asset returns are drawn conditionally on the hidden asset according to

$$P(+_i|+_h) = p + (1-p)\sqrt{C}, \quad P(-_i|-_h) = 1 - p + p\sqrt{C}, \quad C \in [0; 1]. \tag{4.56}$$

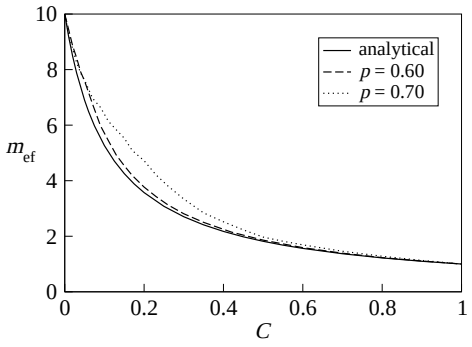


Figure 4.11: The effective size of the Kelly portfolio: a comparison of the approximate result Eq. (4.54) with numerical treatment of Eq. (4.55) for various winning probabilities. The total number of assets is $M = 10$.

It can be shown that $P(+_i) = p$, $P(-_i) = 1 - p$, $C_{ij} = C$; thus the proposed construction satisfies our demands. The distribution $P(\boldsymbol{R})$ is now given by the formula

$$P(\boldsymbol{R}) = P(+_h) \prod_{i=1}^{M} P(R_i|+_h) + P(-_h) \prod_{i=1}^{M} P(R_i|-_h) \tag{4.57}$$

and Eq. (4.55) can be solved, yielding the optimal fraction f^*. Consequently, the definition relation Eq. (4.52) allows us to interpolate m_{ef} for any M, p, C (interpolation is needed because the right side of the definition equation can be numerically computed only for integer m_{ef}). In Fig. 4.11 we compare this result with Eq. (4.54). As can be seen, when the investment return is small ($p = 0.60$ and less), the approximate result performs well. When $p = 0.70$, differences appear for mediocre values of C.

Finally, we use the established framework to investigate the sensitivity of portfolio performance to the estimate of asset correlations. We assume that all assets have pairwise correlation equal C but the investor optimizes the investment assuming a wrong value C'. In Fig. 4.12, the resulting growth rate G^* is shown as a function of C'. As can be seen, underestimation of correlations ($C' < C$) decreases investment performance dramatically—a naive investor supposing zero correlations can even end up with diminishing wealth. By contrast, a similar overestimation of C results in only a mild decrease of G.

Estimates of the effective portfolio size

One can ask whether Eq. (4.53) can be approximated by a simpler formula. Motivated by Eq. (4.54), the natural guess is

$$m_{\text{ef}} = \frac{M}{1 + (M-1)\langle C \rangle}, \tag{4.58}$$

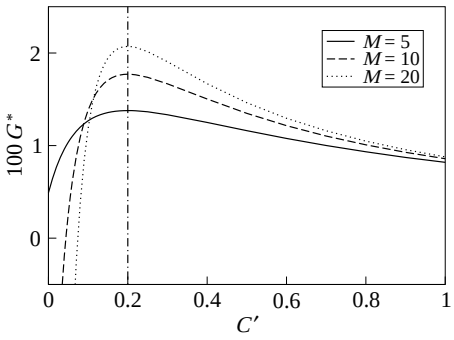


Figure 4.12: The optimal exponential growth rate G^* for an investor assuming a wrong magnitude of the asset correlations. The actual value $C = 0.2$ is marked with the vertical dash-dot line, the winning probability is $p = 0.55$.

we approximate diverse correlations by their average. Since the resulting m_{ef} is the same as Eq. (4.47): this approximation is equivalent to distributing the investment among the assets evenly.

In real markets, assets can be divided into sectors with correlations higher between assets in the same sector than between assets in different sectors. This sector structure can be used to obtain an improved estimate of m_{ef}. Given N sectors, we denote the intra-sector correlation between the assets from sector I by $\tilde{C}_{II}$ and the inter-sector correlation between the assets from sectors I and J by $\tilde{C}_{IJ}$ (for indices labeling sectors we use capital letters). Here $\tilde{C}_{II}$ and $\tilde{C}_{IJ}$ are simple averages ($I, J = 1, \ldots, N$)

$$\tilde{C}_{II} = \frac{1}{M_I^2} \sum_{i,j \in I} C_{ij}, \quad \tilde{C}_{IJ} = \frac{1}{M_I M_J} \sum_{i \in I, j \in J} C_{ij}, \tag{4.59}$$

where M_I is the number of assets in sector I. As a result, an N-dimensional matrix $\tilde{\mathsf{C}}$ is formed. In $\tilde{C}_{II}$ we sum also over diagonal elements of the asset correlation matrix C as it is convenient for our further computation.

Due to the simplifying assumption of identical intra-sector correlations, the optimal investment fractions are identical within a sector and the optimization problem simplifies to N variables f_I. Similarly to Eq. (4.48), the exponential growth rate is now

$$G = \left\langle \ln \left[1 + \sum_{I=1}^{N} f_I \sum_{i \in I} R_i \right] \right\rangle. \tag{4.60}$$

By the same techniques as before, we obtain the estimate of the effective portfolio size

$$m_{\text{ef}} = \sum_{I,J=1}^{M} (\tilde{\mathsf{C}}^{-1})_{I,J}, \tag{4.61}$$

its accuracy will be examined in the following section. Notably, the sector-based estimate of m_{ef} for the Mean-Variance approach is identical.

4.4.3 Correlations in real financial data

Here we test our results on real financial data, keeping two goals in mind. First, to investigate actual values of the effective portfolio size. Second, to examine the accuracy of m_{ef} estimates derived above. We use prices of stocks from the Dow Jones Industrial Average (DJIA) and the Standard & Poor's 500 (S&P 500) which are well-known and common indices consisting of 30 and 500 U.S. companies respectively.

Comparing index and stock variances

Using the daily data for the period 8th April 2004–14th December 2007, we compute daily returns of the DJIA and stocks included by the formula $R(t+1) = (w(t+1) - w(t))/w(t)$, with $w(t)$ denoting the adjusted closing index value on the trading day t. The DJIA value is the sum of the prices of all components, divided by the Dow divisor which changes with time. This effect can be ignored because changes of the divisor are mostly negligible.

During the given period, the variance of the DJIA daily returns was $\sigma^2_{\text{DJIA}} \approx 5.22 \cdot 10^{-5}$, the average variance of the daily returns of the DJIA components was $\sigma^2_{\text{C}} \approx 1.73 \cdot 10^{-4}$. These two figures contradict the assumption of zero correlations because then the variance should scale with the number of assets M as $1/M$ (this follows also from Eq. (4.42) when $\mathsf{C} = \mathsf{I}$ is substituted). By dividing $\sigma^2_{\text{C}}/\sigma^2_{\text{DJIA}}$ we obtain an alternative estimate of the effective number of assets (this estimate was used already in [162]). In our case it is equal to 3.31 which is much less than the total number of stocks in the DJIA.

Alternatively, we can estimate the effective number of assets using the results of our analysis above. From daily returns, asset correlations can be computed by Eq. (4.38), resulting in $\langle C \rangle = 0.322$. Together with the number of stocks $M = 30$, Eq. (4.58) yields $m_{\text{ef}} \approx 2.90$. This is in a good agreement with the value 3.31 obtained by a different reasoning in the previous paragraph.

The effective size

For our analysis we use the same data about the DJIA stocks as in the previous section, for S&P 500 we use the approximately 15 years period period from 2nd January 1992–15th February 2008 and those 338 stocks out of the current 500 which were quoted in the stock exchange during the whole period. As mentioned above, from the daily financial data one can estimate the correlation matrix C. Consequently, the effective portfolio size can be obtained using C^{-1} and Eq. (4.43), Eq. (4.53); it can be approximated using a sector division

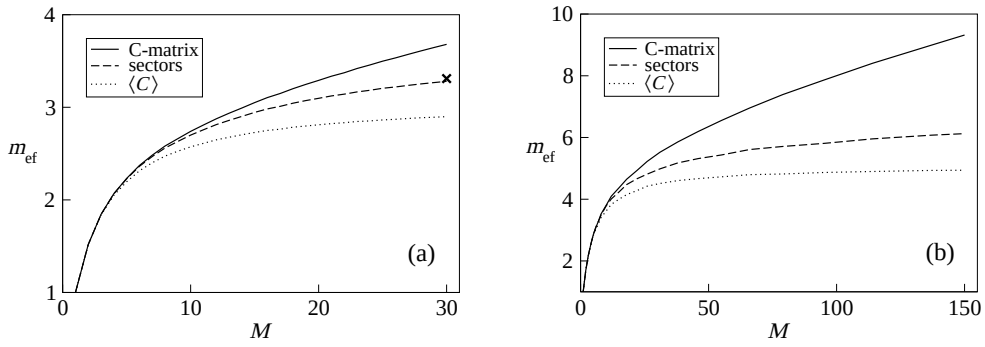


Figure 4.13: The effective size m_{ef} for the DJIA stocks (a) and the S&P 500 stocks (b), computed directly from C, from sector division, and from $\langle C \rangle$. The estimate of the effective size obtained by comparison of variances in Sec. 4.4.3 is shown as a thick cross. Individual realisations fluctuate around the plotted averages with the amplitude approximately one or less.

and Eq. (4.61); or it can be approximated using $\langle C \rangle$ and Eq. (4.58). The stock division into nine sectors is obtained from `http://biz.yahoo.com/p` with the industry sectors: basic materials, conglomerates, consumer goods, finance, healthcare, industrial goods, services, technology, and utilities. Using a different sector division (obtained *i.e.*by minimizing the ratio of intra- and inter-sector correlations) does not influence the results substantially. To obtain the dependency of m_{ef} on the portfolio size M, we select a random subset of M stocks from the complete set and compute m_{ef} for this subset; sensitivity to the subset selection is eliminated by averaging over 5 000 random draws.

The results of the described analysis are shown in Fig. 4.13. For small portfolio sizes (less than approximately ten stocks), the estimates of m_{ef} based on the sector structure or on the average correlation perform well. For a larger portfolio, the sector structure gives a better description of m_{ef} than the average correlation. Nevertheless, both estimates saturate at $M \simeq 20$ while the exact m_{ef} obtained from the complete correlation matrix continues to grow. We see that the effect of heterogeneous correlations increases with M and even the sector structure is then insufficient to describe the system.

The limited horizontal scale in Fig. 4.13 is due to noisy estimates of large correlation matrices from the data with a finite time horizon T. As we are determining $M(M-1)/2$ correlations from MT prices, when T is not very large compared to M, we face underdetermined system of equations [163] (the problem is also known as the curse of dimensionality). It is possible that in Fig. 4.13, the excessive slope of m_{ef} for 5 years data and the complete correlation matrix (in comparison with the slope of the curve for 15 years data) is a signature of this effect. With more frequent financial data, larger portfolios would be easily accessible.

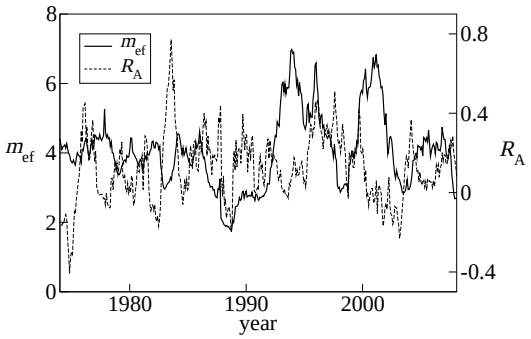


Figure 4.14: Time evolution of the effective size m_{ef} and the average yearly stock return R_A for the DJIA stocks, based on the sliding one year time window.

There is one particular point to be highlighted. According to Eq. (4.47), the estimate of m_{ef} by the average correlation $\langle C \rangle$ is equal to the effective size m'_{ef} of the portfolio containing all available assets with even weights. Such investment was investigated *e.g.*in [164] where it was suggested that the benefits of diversification are exhausted already with a portfolio of ten assets—this result is in agreement with Fig. 4.13. However, since m'_{ef} is lower than the effective sizes obtained directly from a sector division and much lower than those obtained directly from C, we conclude that with an evenly distributed investment one cannot fully exploit the benefits of diversification.

Finally, in Fig. 4.14 we show time evolution of m_{ef} for 20 current stocks from the DJIA.[1] We use the daily data from the period Jan 1973–Apr 2008 and the sliding window with the length one year to obtain estimates of the correlation matrix and compute m_{ef} from C^{-1}. In the same figure we show also the average yearly return of the selected stocks (also estimated on the one-year basis). As we see, m_{ef} varies with a large amplitude, being as low as two at the end of 80's and peaking almost at seven during 1994.

4.4.4 Conclusion

In this work we investigated the influence of correlations on portfolio optimization. This is an important issue—as can be seen in Fig. 4.12, underestimation of correlations can lead to a significant reduction of investment performance. To measure the influence of correlations we suggested a new quantity—the effective portfolio size m_{ef}. Obtained analytical results are accompanied with numerical tests on real financial data. Notably, for most investigated cases, this effective size is much smaller than the actual portfolio size. In particular, evenly distributed investment turns out to be rather ineffective way of diversification.

[1] The selected stocks are AA, BA, CAT, DD, DIS, GE, GM, HON, HPQ, IBM, JNJ, KO, MCD, MMM, MO, MRK, PG, UTX, WMT, and XOM.

Numerical results for m_{ef} could be refined using high frequency financial data which would allow for a shorter time window and lesser averaging artifacts.

Although illustrated only on the Mean-Variance and the Kelly portfolio, the effective size is a general concept which can be used also in other methods for portfolio optimization. Its eventual direct application to portfolio management remains a future challenge.

4.5 Appendices

Here we discuss technical parts of this chapter.

4.5.1 Approximations for the lognormally distributed returns

Our aim is to approximate expressions of the type $\langle g(\eta)\rangle$, where η follows a normal distribution $\varrho(\eta)$ with the mean m and the variance D. For small values of D, this distribution is sharply peaked and an approximate solution can be found expanding $g(\eta)$ around this m. This expansion has the following effective form

$$g(\eta) \overset{\text{ef.}}{=} g(m) + \frac{1}{2}(\eta - m)^2 g^{(2)}(m) + \frac{1}{24}(\eta - m)^4 g^{(4)}(x), \tag{4.62}$$

where $x \in [m, \eta]$. Here we dropped the terms proportional to $(\eta - m)^k$ with an odd exponent k, for they vanish after the averaging. If we take only the first two terms into account, we obtain

$$\big\langle g(\eta)\big\rangle = \int_{-\infty}^{\infty} g(\eta)\varrho(\eta)\,\mathrm{d}\eta \approx g(m) + \frac{D}{2}\, g^{(2)}(m). \tag{4.63}$$

This approximation is valid when the following term of the Taylor series brings a negligible contribution Δ. We can estimate it in the following way ($x \in [m, \eta]$)

$$\Delta = \int_{-\infty}^{\infty} \frac{(\eta - m)^4}{24} g^{(4)}(x)\varrho(\eta)\,\mathrm{d}\eta \lesssim \int_{-\infty}^{\infty} \frac{(\eta - m)^4}{24}\, M\varrho(\eta)\,\mathrm{d}\eta = \frac{MD^2}{8}.$$

Here by M label the maximum of $|g^{(4)}(\eta)|$ in the region $\mathcal{X}$ where $\varrho(\eta)$ differs from zero considerably, *e.g.*$\mathcal{X} = [m - 2D, m + D]$. Since $g(x)$ has no singular points in a wide neighbourhood of m, its fourth derivative is a bounded and well-behaved function. Thus M is finite and Δ vanish when D is small.

In particular, in this work we deal with functions of the form $g(\eta_i) = (\mathrm{e}^{\eta} - 1)/[1 + f(\mathrm{e}^{\eta} - 1)]$. If we use Eq. (4.63) with this $g(\eta)$, approximate $1 + f(\mathrm{e}^{m} - 1)$ in the resulting denominators by 1, e^{m} by 1, and $\mathrm{e}^{m} - 1$ by m, we are left with

$$\big\langle g(\eta)\big\rangle \approx m + D\big(1 - 2f\big)/2. \tag{4.64}$$

We widely use this kind of approximations to obtain the leading terms for the optimal portfolio fractions in section 4.3.

4.5.2 Lognormally distributed returns and correlated asset prices

In section 4.3 we considered only uncorrelated asset prices undergoing the geometric Brownian motion of Eq. (4.22). Obviously, this is an idealised model and real asset prices exhibit various kinds of correlations. In order to treat correlated prices we employ the covariance matrix S to characterise the second moment of the stochastic terms $\langle(\eta_i - m_i)(\eta_j - m_j)\rangle = S_{ij}$. The uncorrelated case can be recovered with the substitution $S_{ij} = \delta_{ij} D_i$.

Again, we would like to approximate the term $\big\langle g(\boldsymbol{\eta})\big\rangle \equiv \int g(\boldsymbol{\eta})\varrho(\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\eta}$. Here $\varrho(\boldsymbol{\eta})$ is the probability distribution of $\boldsymbol{\eta}$ and $g(\boldsymbol{\eta})$ is the function of interest. Notice that the correlations impose the use of vector forms for all the quantities of interest. The Taylor expansion of $g(\boldsymbol{\eta})$ around $\boldsymbol{m}$, Eq. (4.62) in the uncorrelated case, takes the form

$$g(\boldsymbol{\eta}) = g(\boldsymbol{m}) + \nabla g(\boldsymbol{m}) \cdot (\boldsymbol{\eta} - \boldsymbol{m}) + \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{m})^T \mathsf{V}(\boldsymbol{m})(\boldsymbol{\eta} - \boldsymbol{m}) + \ldots .$$

Here $\mathsf{V}(\boldsymbol{m})$ is the matrix of second derivatives of the function $g(\boldsymbol{\eta})$, calculated at the point $\boldsymbol{\eta} = \boldsymbol{m}$. Now we can proceed in the same way as before

$$\begin{aligned}
\big\langle g(\boldsymbol{\eta})\big\rangle \;&\approx\; g(\boldsymbol{m}) \int \varrho(\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\eta} + \sum_{i=1}^{N} \partial_i g(\boldsymbol{m}) \int (\eta_i - m_i)\varrho(\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\eta} + \\
&\quad + \frac{1}{2}\sum_{i,j=1}^{N} V_{ij} \int (\eta_i - m_i)(\eta_j - m_j)\varrho(\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\eta} = \\
&= \; g(\boldsymbol{m}) + \frac{1}{2}\sum_{i,j=1}^{N} S_{ij} V_{ij} = g(\boldsymbol{m}) + \frac{1}{2}\,\mathrm{Tr}(\mathsf{SV}).
\end{aligned} \tag{4.65}$$

In the last line we used the symmetry of S. With given $g(\boldsymbol{\eta})$, $\boldsymbol{m}$ and S, we can now solve the equation $\langle g(\boldsymbol{\eta})\rangle = 0$. In particular, these approximations can be cast into Eq. (4.27) which can then be treated as in the uncorrelated case.

5 Conclusions

Despite decades of research, the science of complexity is still emerging. Various models provided us with good intuition about complex behavior (better to say, about a variety of complex behaviors) but apart from computer simulations we still lack a general tool for complexity.

This thesis is not different. It presents models of complex systems from various areas of human activity and answers several questions. From the obtained results, the simple model of product differentiation based on quality and the concept of the effective portfolio size stand out for their simplicity and generality. Truth be known, although the presented models are highly simplified, some complicated problems remain open. Is it possible to improve the performance of opinion diffusion by a proper generalization? Is it possible to seamlessly combine selection by quality with selection by preferences? How to formalize learning of game properties for real stocks?

Fortunately, open problems not only indicate a lack of past success but also hold out hopes for interesting results in future. Importantly, fruits of efforts in complexity research can be tangible—with a good understanding of complex systems we can make more social experiments with computers and less with real people. This should be the grand vision of the field for the coming years.

Curriculum Vitae

Born in Prešov, Slovakia, on 6th December 1979, married.

Education

- Comenius University, Bratislava, Slovakia 1998—2003
  Undegraduate study of physics with specialization "theoretical and mathematical physics".
- Comenius University, Bratislava, Slovakia 2003—2006
  PhD study in statistical physics with Dr. František Slanina as a supervisor.
- Université de Fribourg, Fribourg, Switzerland 2006—present
  PhD study in statistical physics with prof. Yi-Cheng Zhang as a supervisor.

Experience

- Marie Curie Training Site Program, Bergen, Norway 2004
  Three months stay, work with prof. László Csernai.
- FKS (Physics Correspondence Competition) 1998—2005
  Leader, lecturer, problems-preparing-team member.
- STROM (Math Correspondence Competition) 1998—2002
  Organizer, lecturer, fundraiser.

Skills

Experienced in Mathematica, programming languages C and Pascal, TeX, LaTeX, Metapost, and Corel DRAW. Fluent in English, Czech and Slovak (native language). Passive knowledge of Russian, German, and French language.

Selected Publications

- M. Medo, J. Smrek, *Heterogeneous network with distance dependent connectivity*, EPJ B **63**, 2008
- M. Medo, Y.-C. Zhang, *Market model with heterogeneous buyers*, Physica A **387**, 2008
- Y.-C. Zhang, M. Medo, J. Ren, T. Zhou, T. Li, F. Yang, *Recommendation model based on opinion diffusion*, EPL **80**, 2007

- T. Zhou, J. Ren, M. Medo, Y.-C. Zhang, *Bipartite network projection and personal recommendation*, Phys. Rev. E **76**, 2007
- M. Medo, *Distance-dependent connectivity: Yet Another Approach to the Small World Phenomenon*, Physica A **360/2**, 2006
- F. Kardoš and M. Medo, *Impact of throttle on the surface temperatures of pneumatic elastic elements*, Acta Mechanica Slovaca **3-B**, 2005
- M. Medo, *Bayesian Statistics and its applications in HEP*, diploma thesis, Comenius University, 2002

Present address

Home address: Rte de Villars 23, CH-1700 Fribourg, Switzerland
E-mail address: matus.medo@unifr.ch

Declaration of originality

Hereby I declare that the results and concepts presented in this thesis come from no other sources other than my own work. I acknowledge the contribution of my collaborators and colleagues.

Fribourg, 29. 7. 2008 Matúš Medo

Bibliography

[1] G. Nicolis, I. Prigogine, *Exploring complexity: An introduction*, W. H. Freeman, 1989

[2] E. N. Lorenz, *Journal of the Atmospheric Sciences* **20**, 130–141, 1963

[3] J. Gleick, *Chaos: Making a New Science*, Penguin, 1988

[4] E. N. Lorenz, *The Essence of Chaos*, University of Washington Press, 1996

[5] M. M. Waldrop, *Complexity: The emerging science at the edge of order and chaos*, Simon & Schuster, 1993

[6] M. Gardner, *Scientific American* **223**, 120–123, 1970

[7] E. Bonabeau, M. Dorigo, G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*, Oxford University Press, 1999

[8] G. Weng *et al*, *Science* **284**, 92–96, 1999

[9] Ch. Koch, G. Laurent, *Science* **284**, 96–98, 1999

[10] R. Pastor-Satorras, A. Vespignani, *Physical Review Letters* **86**, 3200–3203, 2001

[11] P. W. Anderson, D. Pines, *The Economy As an Evolving Complex System*, Perseus Books, 1988

[12] P. Ball, *Critical Mass: How One Thing Leads to Another*, Farrar, Straus and Giroux, 2004

[13] M. Schroeder, *Fractals, chaos, power laws*, W. H. Freeman, 1991

[14] P. Bak, *How nature works*, Copernicus, 1996

[15] D. Sornette, *Critical phenomena in natural sciences (2nd edition)*, Springer, 2006

[16] J. M. Ottino, *AIChE Journal* **49**, 292–299, 2003

[17] H. A. Simon, *Proceedings of the American Philosophical Society* **106**, 467–482, 1962

[18] D. Helbing, Ed., *Managing complexity: Insights, concepts, applications*, Springer, 2008

[19] P. Erdös, A. Rényi, *Publicationes Mathematicae* **6**, 290–297, 1959

[20] D. J. Watts, S. H. Strogatz, *Nature* **393**, 440–442, 1998

[21] A. L. Barabási, R. Albert, *Science* **286**, 509–512, 1999

[22] R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47–97, 2002

[23] S. N. Dorogovtsev, J. F. F. Mendes, *Advances In Physics* **51**, 1079–1187, 2002

[24] M. E. J. Newman, *SIAM Review* **45**, 167–256, 2003

[25] S. Boccaletti *et al*, *Physics Reports* **424**, 175–308, 2006

[26] L. da F. Costa *et al*, *Advances in Physics* **56**, 167–242, 2007

[27] S. H. Strogatz, *Nature* **410**, 268–276, 2001

[28] M. E. J. Newman, *Contemporary Physics* **46**, 323–351, 2005

[29] J. D. Farmer, J. Geanakopolos, *Power laws in economics and elsewhere*, SFI Technical report, 2006

[30] A. Clauset, C. R. Shalizi, M. E. J. Newman, *Power-law distributions in empirical data*, arXiv: 0706.1062

[31] J.-P. Bouchaud, M. Potters, *Theory of Financial Risk and Derivative Pricing*, Cambridge University Press, 2008

[32] M. Mitzenmacher, *Internet Mathematics* **1**, 226–251, 2004

[33] E. T. Jaynes, *Physical Review* **106**, 620–630, 1957

[34] F. A. Bais, J. D. Farmer, *The physics of information*, SFI Working paper, ITFA-2007-37, 2007

[35] C. Shannon, *Bell System Technical Journal* **27**, 379–423, 1948

[36] P. D. Grünwald, *The Minimum Description Length Principle*, The MIT Press, 2007

[37] E. T. Jaynes, *Probability theory: the logic of science*, Cambridge University Press, 2003

[38] M. Faloutsos, P. Faloutsos, C. Faloutsos, *Comput. Comm. Rev.* **29**, 251–262, 1999

[39] A. Broder *et al*, *Computer Networks* **33**, 309–320, 2000

[40] S. Brin, L. Page, *Computer Networks and ISDN Systems* **30**, 107–117, 1998

[41] J. M. Kleinberg, *J. ACM* **46**, 604–632, 1999

[42] M. McLuhan, *The Gutenberg Galaxy: The Making of Typographic Man*, University of Toronto Press, 1962

[43] A. Ansari, S. Essegaier, R. Kohli, *Journal of Marketing Research* **37**, 363–375, 2000

[44] Y. P. Ying, F. Feinberg, M. Wedel, *Journal of Marketing Research* **43**, 355–365, 2006

[45] R. Kumar *et al*, *Journal of Computer and System Sciences* **63**, 42–61, 2001

[46] N. J. Belkin, *Communications of the ACM* **43**, 58–61, 2000

[47] M. Montaner, B. López, J. L. De La Rosa, *Artificial Intelligence Review* **19**, 285–330, 2003

[48] J. L. Herlocker *et al*, *ACM Transactions on Information Systems* **22**, 5–53, 2004

[49] P. Laureti, L. Moret, Y.-C. Zhang, *Europhysics Letters* **75**, 1006–1012, 2006

[50] Y.-K. Yu, Y.-C. Zhang, P. Laureti, L.Moret, *Physica A* **371**, 732–744, 2006

[51] F. E. Walter, S. Battiston, F. Schweitzer, in *Autonomous Agents and Multi-Agent Systems* (Ed. D. Helbing), Springer, 2008

[52] C. Cattuto, V. Loreto, L. Pietronero, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1461–1464, 2007

[53] P. Resnick, H. R. Varian, *Communications of the ACM* **40**, 56–58, 1997

[54] M. Balabanović, Y. Shoham, *Communications of the ACM* **40**, 66–72, 1997

[55] M. J. Pazzani, *Artificial Intelligence Review* **13**, 393–408, 1999

[56] B. M. Sarwar *et al*, *Application of dimensionality reduction in recommender system–a case study*, ACM WebKDD 2000 Web Mining for E-Commerce Workshop, 2000

[57] S. Maslov, Y.-C. Zhang, *Phys. Rev. Lett.* **87**, 248701, 2001

[58] T. Hofmann, *ACM Transactions on Information Systems* **22**, 89–15, 2004

[59] G. Takacs *et al*, in *Proceedings of KDD Cup and Workshop*, 2007, On the gravity recommendation system

[60] G. Adomavicius, A. Tuzhilin, *IEEE Trans. on Knowledge and Data Engineering* **17**, 734–749, 2005

[61] C. N. Ziegler *et al*, in *Proceedings of the 14th international conference on World Wide Web*, 2005, Improving recommendation lists through topic diversification

[62] S. M. McNee, J. Riedl, J. A. Konstan, in *Conference on Human Factors in Computing Systems*, 2006, Being accurate is not enough: how accuracy metrics have hurt recommender systems

[63] T. Zhou *et al*, *Europhysics Letters* **81**, 58004, 2008

[64] T. Zhou *et al*, *Phys. Rev. E* **76**, 046115, 2007

[65] Y.-C. Zhang, M. Blattner, Y.-K. Yu, *Phys. Rev. Lett.* **99**, 154301, 2007

[66] I. Percival, *Quantum State Diffusion*, Cambridge University Press, 1999

[67] B. J. Kim, H. Hong, M. Y. Choi, *Phys. Rev. B* **68**, 014304, 2003

[68] Q. Ou *et al*, *Phys. Rev. E* **75**, 2007, 021102

[69] R. A. Penrose, *Proc.Cambridge Philos. Soc.* **51**, 406–413, 1955

[70] M. Blattner, P. Laureti, A. Huntziker, *When are recommender systems useful?*, arXiv: 0709.2562

[71] M. Blattner, Y.-C. Zhang, S. Maslov, *Physica A* **373**, 753–758, 2007

[72] S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, 1994

[73] H. Jeong *et al*, *Nature* **407**, 651–654, 2000

[74] M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 404–409, 2001

[75] M. E. J. Newman, *Phys. Rev. E* **64**, 016131, 2001

[76] W. Souma *et al*, *Physica A* **324**, 396–401, 2003

[77] J . Guillaume, *Information Processing Letters* **90**, 215–221, 2003

[78] J. J. Ramasco, S. N. Dorogovtsev, R. Pastor-Satorras, *Phys. Rev. E* **70**, 036106, 2004

[79] R. Lambiotte, M. Ausloos, *Phys. Rev. E* **72**, 066107, 2005

[80] P. Heymann *et al*, in *Proceedings of the international conference on Web search and web data mining*, 2008, Can social bookmarking improve web search?

[81] J. W. Grossman, P. D. F. Ion, *Congressus Numerantium* **108**, 129–131, 1995

[82] A.-L. Barabási *et al*, *Physica A* **311**, 590–614, 2002

[83] T. Zhou *et al*, *Int. J. Mod. Phys. C* **18**, 297–314, 2007

[84] M. E. J. Newman, *Phys. Rev. E* **70**, 056131, 2004

[85] J. J. Ramasco, S. A. Morris, *Phys. Rev. E* **73**, 016122, 2006

[86] M. Li *et al*, *Physica A* **375**, 355–364, 2007

[87] M. Li *et al*, *Physica A* **350**, 643–656, 2005

[88] M. E. J. Newman, *Phys. Rev. E* **64**, 016132, 2001

[89] M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5200–5205, 2004

[90] J. A. Konstan *et al*, *Communications of the ACM* **40**, 77–87, 1997

[91] T. Zhou *et al*, *Ultra accurate personal recommendation via eliminating redundant correlations*, arXiv:0805.4127

[92] J. Reason, *Human error*, Cambridge University Press, 1990

[93] D. Kahneman, A. Tversky, *Prospect theory: an analysis of decision under risk*, Bradford Books, 2004

[94] H. A. Simon, *The Quarterly Journal of Economics* **69**, 99–118, 1955

[95] M. Rabin, *Journal of Economic Literature* **36**, 11–46, 1998

[96] D. McFadden, *Journal of Business* **53**, 13–29, 1980

[97] I. S. Currim, *Journal of Marketing Research* **19**, 208–222, 1982

[98] S. P. Anderson, A. de Palma, *Oxford Economic Papers* **44**, 51–67, 1992

[99] F. Bagnoli *et al*, *Physica A* **332**, 509–518, 2004

[100] G. Akerlof, *Quarterly Journal of Economics* **84**, 488–500, 1970

[101] H. S. Houthakker, *The Review of Economic Studies* **19**, 155–164, 1952

[102] C. Shapiro, *The Bell Journal of Economics* **13**, 20–35, 1982

[103] Y.-C. Zhang, *Physica A* **299**, 104–120, 2001

[104] M. Motta, *The Journal of Industrial Economics* **41**, 113–131, 1993

[105] J. P. Johnson, D. P. Myatt, *The American Economic Review* **93**, 748–774, 2003

[106] P. Champsaur, J.-Ch. Rochet, *Econometrica* **57**, 533–557, 1989

[107] S. Berry, J. Waldfogel, *Product Quality and Market Size*, NBER Working Paper No. 9675, 2003

[108] J. Sutton, *Sunk Costs and Market Structure*, MIT Press, 1991

[109] D. W. Carlton, J. M. Perloff, *Modern Industrial Organization (4th edition)*, Addison-Wesley, 2005

[110] R. Schmalensee, in *New Developments in the Analysis of Market Structure* (Ed. J. E. Stiglitz, G. F. Mathewson), MIT Press, 1986

[111] Y.-C. Zhang, *Physica A* **350**, 500–532, 2005

[112] R. P. McAfee, *Economics: The Open-Access, Open-Assessment E-Journal* **1**, 1, 2007

[113] M. Mussa, S. Rosen, *Journal of Economic Theory* **18**, 301–317, 1978

[114] J. Stiglitz, in *Prix Nobel: The Nobel Prizes 2001* (Ed. T. Frangsmyr), The Nobel Foundation, 2002

[115] W. B. Arthur, S. N. Durlauf, D. A. Lane (Eds.), *The Economy as An Evolving Complex System II*, Addison-Wesley, 1997

[116] D. Challet, M. Marsili, Y.-C. Zhang, *Minority games: interacting agents in financial markets*, Oxford University Press, 2004

[117] A. De Martino, M. Marsili, *Journal of Physics A* **39**, 465–540, 2006

[118] J. Y. Bakos, *Management Science* **43**, 1676–1692, 1997

[119] M.-J. Oméro *et al*, *J. Phys. I France* **7**, 1723–1732, 1997

[120] A. Lage-Castellanos, R. Mulet, *Physica A* **364**, 389–402, 2006

[121] M. Medo, Y.-C. Zhang, *Physica A* **387**, 2889–2908, 2008

[122] F. Lindskog, *Master Thesis*, ETH Zurich, 2000 Modelling dependence with copulas and applications to risk management

[123] R. B. Nelsen, *An introduction to copulas*, Springer-Verlag, 1999

[124] H. M. Markowitz, *The Journal of Finance* **7**, 77–91, 1952

[125] W. F. Sharpe, *The Journal of Finance* **19**, 425–442, 1964

[126] A. F. Perold, *Management Science* **30**, 1143–1160, 1984

[127] H. Konno, H. Yamazaki, *Management Science* **37**, 519–531, 1991

[128] D. Duffie, *Dynamic Asset Pricing Theory (3rd Edition)*, Princeton University Press, 2001

[129] E. J. Elton, M. J. Gruber, S. J. Brown, W. N. Goetzmann, *Modern Portfolio Theory and Investment Analysis (7th Edition)*, Wiley, 2006

[130] J. L. Kelly, *IEEE Transactions on Information Theory* **2**, 185–189, 1956

[131] L. Breiman, *Fourth Berkeley Symp. Math. Stat. Prob.* **1**, 65–78, 1961

[132] M. Finkelstein, R. Whitley, *Advances in Applied Probability* **13**, 415–428, 1981

[133] S. Browne, in *Finding the Edge, Mathematical Analysis of Casino Games* (Ed. O. Vancura, J. Cornelius, W. R. Eadington), University of Nevada, 2000

[134] E. O. Thorp, in *Finding the Edge, Mathematical Analysis of Casino Games* (Ed. O. Vancura, J. Cornelius, W. R. Eadington), University of Nevada, 2000

[135] H. A. Latané, *The Journal of Political Economy* **67**, 144–155, 1959

[136] P. A. Samuelson, *PNAS* **68**, 2493–2496, 1971

[137] H. Levy, *International Economic Review* **14**, 601–614, 1973

[138] H. M. Markowitz, *The Journal of Finance* **31**, 1273–1286, 1976

[139] M. Marsili, S. Maslov, Y.-C. Zhang, *Physica A* **253**, 403–418, 1998

[140] S. Maslov, Y.-C. Zhang, *International Journal of Theoretical and Applied Finance* **1**, 377–387, 1998

[141] T. M. Cover, *Mathematical Finance* **1**, 1–29, 1991

[142] C. Whitrow, *Appl. Statist.* **56**, 607–623, 2007

[143] S. J. Grossman, J. E. Stiglitz, *The American Economic Review* **70**, 393–408, 1980

[144] A. Capocci, Y.-C. Zhang, *International Journal of Theoretical and Applied Finance* **3**, 511–522, 2000

[145] S. Browne, W. Whitt, *Adv. Appl. Prob.* **28**, 1145–1176, 1996

[146] D. Sivia, J. Skilling, *Data Analysis: A Bayesian Tutorial*, Oxford University Press, 2006

[147] M. Medo, Y. M. Pis'mak, Y.-C. Zhang, *Novel variations of the Kelly game*, arXiv:0803.1364

[148] F. Slanina, *Physica A* **269**, 554–563, 1999

[149] S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi, *Science* **220**, 671–680, 1983

[150] V. Černý, *Journal of Optimization Theory and Applications* **45**, 41–51, 1985

[151] R. C. Merton, *The Review of Economics and Statistics* **51**, 247–257, 1969

[152] C. W. Gardiner, *Handbook of Stochastic Methods (3rd Edition)*, Springer, 2004

[153] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry (3rd Edition)*, North Holland, 2007

[154] H. A. David, H. N. Nagaraja, *Order Statistics (3rd Edition)*, Wiley, 2003

[155] J. Lintner, *The Journal of Finance* **20**, 587–615, 1965

[156] P. Jorion, *The Journal of Business* **58**, 259–278, 1985

[157] M. Statman, *The Journal of Financial and Quantitative Analysis* **22**, 353–363, 1987

[158] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, Wiley, 1991

[159] S. L. Heston, K. G. Rouwenhorst, *Journal of Financial Economics* **36**, 3–27, 1994

[160] V. Polkovnichenko, *Review of Financial Studies* **18**, 1467–1502, 2005

[161] R. C. Merton, *The Journal of Financial and Quantitative Analysis* **7**, 1851–1872, 1972

[162] W. N. Goetzmann, A. Kumar, *Why do individual investors hold under-diversified portfolios?*, working paper, 2004

[163] L. Laloux *et al*, *Phys. Rev. Lett.* **83**, 1467–1470, 1999

[164] E. J. Elton, M. J. Gruber, *The Journal of Business* **50**, 415–437, 1977